ITC Kenya Wave 2

Technical Report

October 2020

ITC Kenya Research Team
Dr. Lawrence Ikamari (Co-Principal Investigator, University of Nairobi, Population Studies and Research Institute [PSRI])
Dr. Jane Rahedi Ong'ang'o (Co-Principal Investigator, Kenya Medical Research Institute [KEMRI]
Dr. Andrew Mutuku (Project Manager, University of Nairobi, PSRI)
Ms. Emma Wanyonyi (Former CEO, International Institute for Legislative Affairs [IILA])

ITC Kenya International Research Team (University of Waterloo, Canada)
Professor Geoffrey T. Fong (Principal Investigator, Department of Psychology)
Professor Mary E. Thompson (Principal Investigator, Department of Statistics and Actuarial Science)
Dr. Christian Boudreau (Co-Investigator, Department of Statistics and Actuarial Science)
Dr. Anne C.K. Quah (Senior Research Scientist, Department of Psychology)
Dr. Susan Kaai (Research Scientist, Project Manager, Department of Psychology)
Ms. Yingchen Fan (Data Analyst, Department of Psychology)
Dr. Pete Driezen (Senior Data Analyst, Department of Psychology)

Funding
Canadian Institutes of Health Research Foundation Grant (FDN-148477)

Ethics Statement
The survey protocols and all materials, including the survey questionnaires, were cleared for ethics by Office of Research Ethics, University of Waterloo, Canada (ORE#30136). All participants provided consent to participate.

Suggested Citation: ITC Project. (2020, October). ITC Kenya Wave 2 (2018) Technical Report. University of Waterloo, Waterloo, Ontario, Canada; University of Nairobi, Ministry of Health, and Kenya Medical Research Institute, Kenya.

Table of Contents

List of Figures and Tables

Preface to ITC Kenya Wave 2 Technical Report

This report documents the second wave of the International Tobacco Control Policy Evaluation (ITC) Survey that was conducted in Kenya. The ITC Kenya Wave 1 Survey was conducted from October to December 2012. The second wave was conducted from April to June 2018, i.e., about five years after the first wave.

For the most part, the format of this report is similar to the Wave 1 technical report. However, there were a number of changes to some content and methods in the second wave:

1. Respondents from the first wave were recontacted to participate in the second wave.
2. New respondents were recruited to replace those lost to attrition, using the same sampling strategy to the one used in Wave 1.
3. Prefilled Household Recontact Forms and Cover Pages were prepared based on Wave 1 information.
4. Two new screeners (Screeners 2 and 3) were added at Wave 2.
5. Prefilled Recontact Labels were designed to eliminate the need to manually fill the series of ID Codes (for Wave 1 Recontact Respondents) in the cover pages of the Screeners and the Individual Surveys.
6. An updated description of cross-sectional and longitudinal weights for Waves 1 and 2 is included in this report.

1 Introduction

1.1 Background

The International Tobacco Control Policy Evaluation (ITC) Project is a multi-country prospective cohort study designed to measure the psychosocial and behavioral impact of key policies of the World Health Organization Framework Convention on Tobacco Control (WHO FCTC). It is a transdisciplinary collaboration of over 150 researchers across 29 countries—Canada, United States, United Kingdom, Australia, Ireland, Thailand, Malaysia, Republic of Korea, China, Mexico, Uruguay, New Zealand, France, Germany, The Netherlands, Mauritius, Brazil, Bangladesh, Bhutan, India, Kenya, Zambia, Greece, Hungary, Romania, Poland, Spain, Japan, and Abu Dhabi-United Arab Emirates. The ITC Project now involves countries inhabited by over 50% of the world's population, over 60% of the world's smokers, and over 70% of the world's tobacco users. The ITC Project is led by Professor Geoffrey T. Fong at the University of Waterloo. In 2012, two African countries (Zambia and Kenya) joined the ITC Project and have completed two waves of data collection.

On June 25, 2004, Kenya ratified the Framework Convention on Tobacco Control (FCTC):

- In 2004, smoking was banned in restaurants, healthcare and educational facilities, universities, government facilities, public transit, indoor offices, other indoor workplaces, bars and pubs.
- In August 2007, the Kenyan government passed the Tobacco Control Act, which was assented into law on September 27, 2007. The law came into force in July 2008 and it included:
    - Display of 13 rotating text-only warnings on 30% of the front and 50% of the back of the package in Kiswahili and English.
    - Ban on misleading/deceptive packaging but does not specifically ban terms such as "light", "mild", and "low tar".
    - The sale of 10 or fewer cigarettes was prohibited.
    - Ban on all advertisement of tobacco products on any medium of communication.
    - Ban on all sponsorship and smoking in public places except for designated smoking areas.
    - Ban on the sale of tobacco products to youth under 18.
    - Required the disclosure of tar, nicotine, and other toxic constituents. It is limited to the disclosure of contents and not their quantities.
    - Ban on the sale of tobacco products in automatic vending machines.
- Between 2007 and 2009, the government increased taxation of tobacco products to 5% of retail price. In 2011, the tax system on cigarettes was simplified, i.e., changed from tiered specific tax to an ad valorem tax.
- Additionally, Kenya developed the National Tobacco Control Action Plan (NTCAP) 2010-2015, which outlines the key directions for tobacco control in Kenya over a five-year period.
- In December 05, 2014, the following new "Tobacco Control Regulations, 2014" were tabled in parliament to be implemented six months later (June 05, 2015):
    - 15 new pictorial health warnings (PHWs) that cover 30% of the front and 50% of the back of the pack.
    - Ban on smoking in cars with children.

- Ban on smoking on streets/walkways near public places.
- Certification of compliance required for designated smoking areas (DSAs).
- Requirement for manufacturers and importers to disclose product ingredients and sales/production data.
- Restriction on public authority-tobacco industry interactions.
- Measures to prevent industry interference.
- Requirement for manufacturers or importers to pay 2% of value of products manufactured or imported to Tobacco Control Fund annually.

Enforcement of these tobacco control laws is weak due to tobacco industry interference. The laws were implemented in July 05, 2015 but were suspended in High Court in July 02, 2015 by British American Tobacco-Kenya (BAT-K). However, only three out of 15 PHWs were released in the Kenya retail market in September 26, 2016 (https://www.the-star.co.ke/news/big-read/2016-09-26-this-is-how-your-cigarette-packet-should-look-like-beginning-today/). There was another legal challenge by BAT-K on October 18, 2016, where the High Court ruled in favor of the government. In February 2017, BAT-K challenged the health warnings provisions in the 2014 Tobacco Control Regulations. The tobacco industry was finally defeated in November 2019 and the 2014 Regulations were upheld and the three PHWs are still appearing on the packs.

- In 2015: the *Excise Duty Act* simplified the tobacco tax system to a uniform specific excise tax structure.
- In 2017: a tiered tobacco tax system was reintroduced with different tax rates for filtered and unfiltered cigarettes.
- In December 2017: a ban was introduced on the import, manufacture, sale, use, advertising and promotion, and distribution of Shisha under the *Public Health (Control of Shisha Smoking) Rules, 2017*. The Tobacco Industry went to court but were defeated when the high court upheld the regulation on 26th July 2018 (court documents: http://kenyalaw.org/caselaw/cases/view/156798/).

1.2 Main Objectives

The objectives of the ITC Kenya Survey are:

1) **To examine the patterns of tobacco use behaviour in Kenya. The survey will also provide information about tobacco users' knowledge, beliefs, attitudes and opinions about using tobacco.**
   The ITC Kenya Survey provides patterns of tobacco use among the Kenyan population. It describes the population's consumption patterns and quitting behaviour, as well as its knowledge, beliefs, and attitudes about tobacco use. It also investigates the population's shift from traditional tobacco products (bidis, kreteks, and smokeless) to cigarettes and electronic cigarettes.

2) **To examine the impact of specific tobacco control policies that have been, or will be, implemented in Kenya, on tobacco use and tobacco-related behaviour among tobacco users in Kenya.**
   The ITC Kenya Wave 2 (2018) Survey was conducted about six years after Wave 1, during which time the following tobacco control policies were implemented:
   - **2015:** the *Excise Duty Act* simplified the tobacco tax system to a uniform specific excise tax structure

- **September 2016:** the *Tobacco Control Regulations, 2014* were published and were to take effect in six months' time (but were delayed due to a legal challenge from the tobacco industry). Three of the 15 new pictorial health warnings were implemented September 2016 on cigarette packages
- **2017:** tiered tobacco tax system was reintroduced with different tax rates for filtered and unfiltered cigarettes
- **December 2017:** ban on the import, manufacture, sale, use, advertising and promotion, and distribution of Shisha under the *Public Health (Control of Shisha Smoking) Rules, 2017*

Overall, the ITC Kenya Survey evaluates the impact of tobacco control policies in the following areas of the FCTC: Health warning labels and package descriptors; smoke-free legislation; pricing and taxation of tobacco products as well as the prevalence of compensatory behaviours that may offset the impact of taxation (e.g., cheaper purchasing options, smuggling); education and support for cessation; and tobacco advertising and promotion.

Findings from the ITC Kenya Survey will provide a detailed picture of the current tobacco control policy landscape in Kenya, including the beliefs, attitudes and behaviours of tobacco users and non-users, following the implementation of the WHO Framework Convention on Tobacco Control (FCTC) and the new tobacco regulations that were implemented after Wave 1 (2012).

3) **To compare the psychosocial and behavioural effects of national-level tobacco control policies and programs in Kenya with findings from the other 28 ITC countries.**
The ITC Project aims to provide an evidence base to guide policies enacted under the WHO FCTC, and to systematically evaluate the effectiveness of these legislative efforts. All ITC Surveys are developed using the same conceptual framework and methods, and the survey questions are designed to be identical or functionally equivalent in order to allow strong comparisons across ITC countries. The evaluation studies conducted from the ITC Surveys take advantage of natural experiments created when an ITC country implements a policy: changes in policy-relevant variables in that country from pre- to post-policy survey waves are compared to those of other ITC countries where that policy has not changed. This research design provides high levels of internal validity, allowing more confident judgments regarding the possible causal impact of policy.

4) **To suggest changes to current government tobacco policies**
Recommendations to strengthen the current tobacco control policies are made based on existing and derived survey information. The aim is to optimize the effects of tobacco control policies with regard to situational and individual difference moderators: (a) demographic variables; (b) personality variables (e.g., time perspective); (c) environmental context (e.g., number of peers/family members who smoke); and (d) the individual's smoking history (e.g., past quit attempts, smoking intensity, and quitting smoking).

1.3 The Research Team

The ITC Kenyan Survey fieldwork was conducted by team members from the University of Nairobi, Population Studies and Research Institute (PSRI), Kenya Medical Research Institute (KEMRI), and International Institute

for Legislative Affairs (IILA). The research team is collaborating with an international team of researchers at the University of Waterloo, Canada.

1.4 Interview Procedures

Data were collected using 'face-to-face' interviewing methods. The survey questionnaires (referred to as survey herewith) were adapted by the ITC Kenya and Waterloo teams to ensure that they were relevant to the Kenyan context. Participants gave their informed consent before commencing an interview. The surveys took approximately 60 minutes for tobacco users and approximately 30 minutes for non-users to complete. The bilingual (English and Kiswahili) surveys were administered in Kiswahili, which is the Kenyan national language or English (official language) for those respondents who wished to complete the survey in English.

A total of 45 interviewers, 19 Supervisors, 4 Regional Coordinators, 1 Project Manager, 1 Data Manager, and 4 data entry clerks were contracted by University of Nairobi PSRI. Each province had a team that comprised of a supervisor and 2-4 interviewers. Each team remained in their province until data collection was completed. A notable number of the fieldwork team had participated at Wave 1. Data collection commenced in April 09, 2018 and was completed in June 19, 2018.

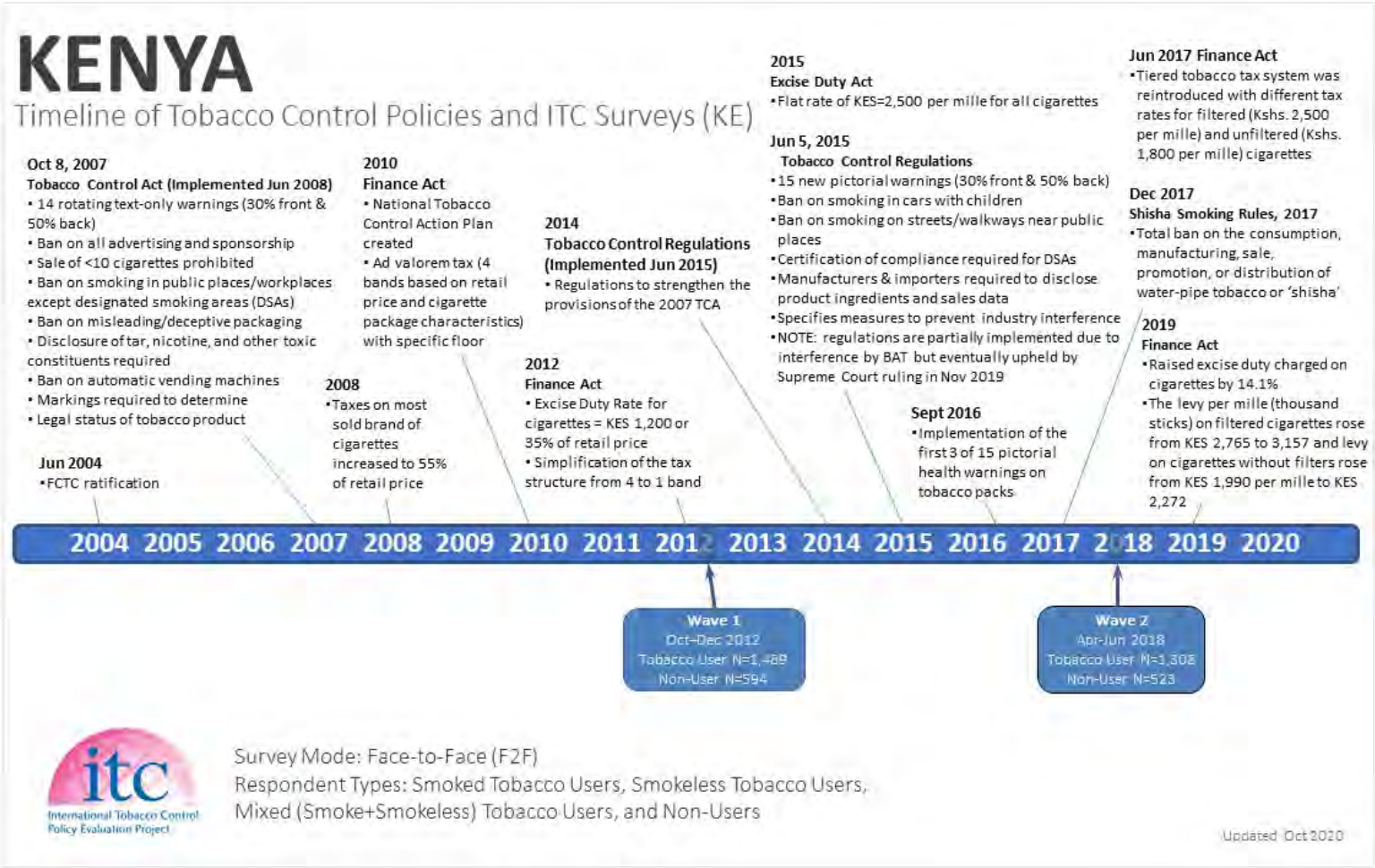
1.5 Pilot Testing of Surveys

A pilot survey was conducted during the interviewers' training workshop on March 12-16, 2018. Interviewers practiced what they had learned in the training workshop. A debriefing session was held after the pilot testing activity. Interviewers shared their experiences regarding the field logistic challenges and the actual interviewing exercise. The surveys were revised based on the de-briefing exercise.

1.6 ITC Kenya Survey Timelines

The ITC Survey is a longitudinal cohort study. Therefore, the respondents who participated in Wave 1 were re-contacted at Wave 2 to answer the follow-up survey. For respondents that could not be recontacted, the sample was replenished to retain the target sample size of 1,500 tobacco users and 600 non-users of tobacco. Figure 1 shows the timeline of the ITC Kenya Project.

Figure 1: ITC Kenya Survey Timeline



2 ITC Kenya Wave 2 Sampling Design

2.1 Sampling Design for ITC Kenya Wave 2

The ITC Kenya Survey was designed to be a nationally representative sample of approximately 1,500 tobacco users and 600 non-users of tobacco aged 15 years or older selected through a multi-stage clustered sampling design. Specifically, at Wave 1, the design was stratified by province and sampled a total of 148 clusters/enumeration areas (EA), allocated to the provinces in numbers proportional to population size. Design calculations were based on 2009 Kenya Population Census data. Within each EA, it was intended that 10 tobacco users and 4 non-users of tobacco would be interviewed. See Appendix A for details of how the original sampling of the ITC Kenya Wave 1 Survey was conducted.

At Wave 2, every effort was made to recontact Wave 1 households; dropouts were replaced by adults of the same tobacco use status in newly enumerated households, in the same EA where possible, sampled using the same procedure as in Wave 1. For Wave 2 updated maps were used for the replenishment sample. It was intended that if an EA (cluster) was exhausted, an additional EA (cluster) would be randomly drawn from the same sub-location (See Table 1).

2.2 Recontact Sampling for ITC Kenya Wave 2

The Wave 2 sampling plan focused on recontacting as many Wave 1 respondents as possible, even if they were smokers at Wave 1 but had quit smoking in Wave 2. There was a separate quitter survey for those Wave 1 respondents who had quit smoking. In order to minimize ID code errors at Wave 2, field interviewers were given Household Recontact Forms (HRFs) with prefilled ID numbers. The Statistician from Waterloo used ID Code information from the Wave 1 dataset to pre-fill the Wave 2 HRFs. Additionally, prefilled ID labels were prepared (in duplicates) and were pasted on the cover page of the Recontact Screeners and Recontact individual surveys. The cover page for the Recontact Household Survey was also prefilled to minimize ID Code errors.

2.3 Replenishment Sampling for ITC Kenya Wave 2

The replenishment sample was made up of new respondents who were interviewed to replace Wave 1 respondents who could not be traced at Wave 2 (i.e., lost to follow up). They were to be taken from newly enumerated households from the same EA where possible. If an EA was exhausted, a new EA was sampled by Waterloo Statisticians and sent to the team.

Table 1: Selected EAs to replace exhausted EAs (Replenishment Sample)

Wave 1 EAs: County (Province)	Wave 1 EA Name and No	A new EA selected to replace the exhausted/inaccessible W1 EA	ID for EA
Nairobi North (Nairobi)	Uhuru Phase 3 EA=05029	Rabai EA=04929	ID: Prov=01 Dist=03 Div=03 Loc=01 Subloc=02
Nyandarua North (Central)	Local 4 'A' Village/Estate –EA No. 00119	Kanyagia Town EA No. 00319	ID: Prov=02 Dist=01 Div=01 Loc=06 Subloc=01
Nyandarua North (Central)	Migunda Mieru Village/Estate –EA No. 00719	Kanyagia Central EA No. 00619	ID: Prov=02 Dist=01 Div=01 Loc=06 Subloc=01
Mombasa (Coast)	Nyali Phase 1– EA No. 11629	Ratna B- EA No. 13929	ID: Prov=03 Dist=01 Div=03 Loc=02 Subloc=01
Garissa (North Eastern)	Windsor Village/Estate - EA No. 00229	Windsor EA No. 00129	ID: Prov=05 Dist=01 Div=01 Loc=01 Subloc=02
Garissa (North Eastern)	Garissa Ndogo 'A' Village/Estate - EA No. 03629	Garissa Ndogo 'A' EA No. 03529	ID: Prov=05 Dist=01 Div=01 Loc=01 Subloc=01
Naivasha (Rift Valley)	Kidong Ranch EA No. 00139	Rift 'C' Quarters EA No 01339	ID: Prov=07 Dist=32 Div=02 Loc=03 Subloc=02
Naivasha (Rift Valley)	Kamere EA No. 05239		ID: Prov=07 Dist=32 Div=02 Loc=03 Subloc=02
Kibwezi (Eastern)	Ilandi/Mukange EA No. 00619	SILIMBILANDI EA No. 00159	ID: Prov=04 Dist=27 Div=01 Loc=03 Subloc=01
Kibwezi (Eastern)	Katangi Village EA No. 01619	ISUNGULUNI EA No. 01519	ID: Prov=04 Dist=27 Div=03 Loc=02 Subloc=02
Nairobi East (Nairobi)	Dandora Phase IV Pline EA=11729	Dandora Phase IV Pline EA=12629	ID: Prov=01 Dist=02 Div=01 Loc=01 Subloc=02
Nairobi East (Nairobi)	Dandora Phase IV EA=03729	Dandora Phase IV EA=04729	ID: Prov=01 Dist=02 Div=01 Loc=01 Subloc=02
Nairobi East (Nairobi)	D3/Kayole EA=39629	D3/Kayole EA=39729	ID: Prov=01 Dist=02 Div=01 Loc=04 Subloc=01
Nairobi East (Nairobi)	By Pass EA=01129	By Pass EA=01029	ID: Prov=01 Dist=02 Div=01 Loc=07 Subloc=02
Nairobi East (Nairobi)	By Umoja 1 Caltex=08229	By Umoja 1 Caltex=09329	ID: Prov=01 Dist=02 Div=01 Loc=08 Subloc=01
Nairobi North (Nairobi)	Zawadi EA=04224	Zawadi EA=04524	ID: Prov=01 Dist=03 Div=03 Loc=03 Subloc=02
Nairobi North (Nairobi)	Kamande 'B' EA=02929	Kamande 'B' EA=02429	ID: Prov=01 Dist=03 Div=02 Loc=01 Subloc=02
Kilindini (Coast)	Amani 'A' EA=05429	Amani 'A' EA=05329	ID: Prov=03 Dist=02 Div=02 Loc=02 Subloc=01
Mombasa (Coast)	Nyali Phase 1 EA=11629	Ratna 'B' EA=13929	ID: Prov=03 Dist=01 Div=03 Loc=02 Subloc=01
Teso North District (Western)	Kamosenu+Kataboi EA No. 00119	Akobwait EA No. 00919	ID: Prov=08 Dist=17 Div=01 Loc=07 Subloc=01
Teso North District (Western)	Akirimasi + Akobwait EA No. 00219	Ochwa EA No. 00319	ID: Prov=08 Dist=17 Div=01 Loc=06 Subloc=01

Wave 1 EAs: County (Province)	Wave 1 EA Name and No	A new EA selected to replace the exhausted/inaccessible W1 EA	ID for EA
Teso South District (Western)	Amagoro+Okisimo EA No. 00319	Komelo EA No. 00219	ID: Prov=08 Dist=20 Div=01 Loc=06 Subloc=01
Lugari (Western)	Bahai EA=00219	Lukova EA=00319	ID: Prov=08 Dist=05 Div=01 Loc=01 Subloc=02
Lugari (Western)	AP Camp+Baharini EA=01119	Mukuva EA=00819	ID: Prov=08 Dist=05 Div=02 Loc=03 Subloc=01
Kericho (RV)	Tabet (Simisbei/Tabet/ Sumuiyot) EA=00219	Kamaas EA=00319	ID: Prov=07 Dist=39 Div=01 Loc=04 Subloc=02
Naivasha (RV)	Kihoto 'B' EA=03029	Kihoto 'B' EA=02929	ID: Prov=07 Dist=32 Div=02 Loc=01 Subloc=01
Migori (Nyanza)	Nyamasongo+Nyaranga EA=00819	Mkoro Bamgot EA=01519	ID: Prov=06 Dist=10 Div=02 Loc=01 Subloc=01
Migori (Nyanza)	Nyamanga 'B' EA=00319	Misare 'B' EA=01619	ID: Prov=06 Dist=10 Div=02 Loc=01 Subloc=02

Interviewers were given maps of the EAs that were drawn at Wave 1. If there were changes in the dwellings in the EA, they first updated the Wave 1 maps. New dwellings were allocated numbers that were not used in Wave 1 in a clockwise manner. Interviewers used the Wave 1 "Random Tables" (See Appendix B1) to randomly select the dwellings for the new households. Specifically, they started with the dwelling in the first row below where they left off in Wave 1; then they went to the dwelling in the next row below etc., until they completed the replenishment sample. For EAs that had more than 500 households, the Waterloo Statistician sent an updated Random Table which included additional columns to cater for households with more than 500 households (see Appendix B2).

The recontact/replenishment quota target for Wave 2 (per EA) was 10 tobacco users and 4 non-users. At Wave 2, the number of tobacco users and non-users who completed recontact interviews were 678 and 273 respectively. New or replenishment respondents, 630 tobacco users and 250 non-users of tobacco were recruited at Wave 2 (See Table 2).

Table 2. Number of Recontact and Replenishment Respondents at Wave 2

	Tobacco Users	Non-Users of Tobacco	Total
Recontact	678	273	**951**
Replenishment	630	250	**880**
Total	**1308**	**523**	**1831**

3 Survey Protocols

3.1 Recontact Respondents Selection and Consent

Respondents who were interviewed in Wave 1 were recontacted in Wave 2. Field interviewers were asked to re-introduce themselves to respondents using a pre-determined script. Details of the field procedure are summarized in a flow chart in Appendix C.

Once contact with a respondent from Wave 1 was re-established, the interviewer first allowed the respondent to read (if they were literate) the information letter about the research study before completing the consent form, i.e., the respondent signed (or thumb stamped if the respondent could not sign) two copies of the consent form.

If a respondent could not read, the interviewer orally explained the study details to the respondent using the information letter as a guide before requesting a signature or thumb stamp. The interviewer gave the respondent the information letter and a copy of the signed (or thumb stamped if the respondent could not sign) consent form. The second copy of the consent form and the screener (see Section 3.3) were attached to the completed survey for that respondent.

3.2 Replenishment Respondents Selection and Consent

Replenishment was only done in newly enumerated households at Wave 2 (within the same EAs at Wave 1), not in households that had been enumerated in Wave 1. However, as described in Section 2.3, sometimes the EAs were exhausted and the interviewers were given newly sampled EAs by the Waterloo Statistician. Only Kenya citizens aged 15 or older were eligible for replenishment.

3.2.1 Selection of Eligible Household Members

There were two different categories of eligible respondents in a replenishment household:

- Tobacco users – The interviewer could randomly select up to 4 tobacco users from every household (with priority to female tobacco users) as was conducted in Wave 1.
- Non-users – The interviewer could randomly select 1 non-user at every 4th household (Central and Eastern), every 5th household (Nairobi, Coast, Rift Valley, and North-Eastern provinces) or every 7th household (Nyanza and Western provinces) from which a non-user could be selected with gender matching what was required from that household to meet the alternating gender requirement – as was done in Wave 1.

Once a replenishment respondent was selected, the information letter was provided and consent was obtained as explained in Section 3.1.

3.3 Surveys and Screeners

There were three individual screeners (Screeners 1, 2, and 3), one household survey (H), and five individual surveys (T, L, M, N, and Q).

3.3.1 Types of Screeners

The purpose of the screeners was to establish the tobacco status of a respondent at Wave 2.

- **Replenishment Screener 1**: Replenishment Respondent: For new eligible respondents at Wave 2 who were selected (excluding those who exclusively smoke Bidis) to replace Wave 1 respondents who could not be recontacted at Wave 2. This screener was to ensure that the appropriate individual survey (T, L, M, or N) was administered at Wave 2.
- **Recontact Screener 2**: Tobacco User (Recontact): For every Wave 1 respondent who was a tobacco user at Wave 1. This screener was to ensure that the appropriate individual survey (T, L, M, or Q) was administered at Wave 2.
- **Recontact Screener 3**: Non-User of Tobacco (Recontact) - For every Wave 1 respondent who was a non-user of tobacco at Wave 1. This screener was to ensure that the appropriate individual survey (T, L, M, or N) was administered at Wave 2.

3.3.2 Recontact Surveys

These surveys were used for Wave 1 respondents who were successfully recontacted at Wave 2. Below is a brief description of each recontact survey:

- **Recontact Household (H) Survey**: Only for the Wave 1 Head or Key Informant of the Household, i.e., one H Survey per household.
- **Recontact Smoked Tobacco (T) Survey**: For Wave 1 respondents who smoked cigarettes and/or bidis at least once a month.
- **Recontact Smokeless Tobacco user (L) Survey:** For Wave 1 respondents who used smokeless tobacco at least once a month.
- **Recontact Mixed Tobacco user (M) Survey:** For Wave 1 respondents who used BOTH smoked tobacco (cigarettes and/or bidis*) and smokeless tobacco at least once a month.
- **Recontact Non-User of Tobacco (N) Survey**: For Wave 1 respondents who did not smoke smoked tobacco (cigarettes and/or bidis) or use any smokeless tobacco products at least once a month.
- **Recontact Quitter (Q) Survey**: For Wave 1 respondents who were tobacco users in Wave 1 but had completely quit tobacco (cigarettes/bidis*, or/and smokeless tobacco) at Wave 2.

* Due to the extremely low prevalence of bidis smoking, it was decided that bidis smokers from Wave 1 would be recontacted but that no bidis smokers would be recruited in the replenishment sample.

3.3.3 Replenishment Surveys

These surveys were for new respondents at Wave 2 and for non-user respondents from Wave 1 who started smoking at Wave 2. Below is a brief description of each replenishment survey:

- **Household (H) Survey**: Only for the Wave 2 Head or Key Informant of the Household, i.e., one H survey per household.
- **Smoked Tobacco (T) Survey**: For Wave 2 respondents who smoked cigarettes at least once a month.
- **Smokeless Tobacco user (L) Survey:** For Wave 2 respondents who used smokeless tobacco at least once a month.

- **Mixed Tobacco user (M) Survey:** For Wave 2 respondents who used BOTH cigarettes and smokeless tobacco at least once a month.
- **Non-user of tobacco (N) Survey**: For Wave 2 respondents who did not smoke cigarettes or use any smokeless tobacco products (N) at least once a month.

3.4 Survey Content

The research design focuses on how individuals respond to policies and how they change over time. Below is a general description of the main constructs assessed in ITC surveys:

- Demographic variables: These include questions to assess gender, age, ethnicity, education, number of smokers in the household; smoker's state of health, religion, socioeconomic status.
- Proximal variables: These include measures assessing awareness of a policy (e.g., of warning labels, cessation assistance, advertising and promotion) and, where relevant, cognitive processing as a result of exposure to the policy (e.g., thinking about health warnings).
- Distal variables: Questions assessing distal variables include those that test psychosocial theories (e.g., the theory of planned behavior: attitudes, subjective norms, perceived behavioral control/self-efficacy), risk perceptions, quit intentions, and other relevant measures. In addition, the survey included questions that measure smokers' self-exempting beliefs, that is, those that many smokers hold that may help to sustain their smoking behavior.
- Moderator variables: Questions about moderator variables include items assessing perceived time perspective (i.e., the tendency for individuals to think about the long-term versus short-term consequences of their actions, which is a predictor of smoking behavior) and stress in addition to the demographic variables listed above. It also examines background variables such as country, region, and community size.
- Tobacco use behavior variables: Standard questions assessing tobacco use behaviors developed by the WHO will be utilized. There are questions that measure a variety of aspects of tobacco use behavior including usual brand, quit intentions, and other smoking-relevant constructs.
- Questions enquiring about household income, expenditures, wealth, and tobacco cultivation were included.

3.5 Survey Sections

All surveys had similar characteristics. Table 3 shows a description of these characteristics and the sections that were relevant to each survey. Each survey was divided into a number of sections that were arranged in a specific order as shown on the table. In each survey section, the total number of questions across the different surveys was not necessarily the same.

Table 3: Section Headings per Questionnaire Type

Survey Sections		Survey Types				
		M	T	L	N	Q
SMOKELESS TOBACCO	PAST & PRESENT FREQUENCY	√	√	√	√	√
	WHEN AND WHY	√		√		
	DEPENDENCE	√		√		
	QUITTING ATTEMPTS	√		√		
	PLANS TO QUIT	√		√		
	REASONS TO QUIT	√		√		
	BRAND CHOICE	√		√		
	LAST PURCHASE	√		√		
	KNOWLEDGE	√		√		√
	PSYCHOSOCIAL BELIEFS	√		√		√
	HEALTH WARNINGS	√		√		√
	SHOW PACK SECTION	√		√		
	PERCEIVED RISK	√		√		√
	MINI-QUIT		√		√	√
	REASONS DID QUIT		√		√	√
	STAY QUIT					√
	KNOWLEDGE (NON-USERS)		√		√	√
	PSYCHOSOCIAL BELIEFS (NON-USERS)		√		√	√
	HEALTH WARNINGS (NON-USERS)		√		√	√
CIGARETTE	PAST & PRESENT FREQUENCY	√	√	√	√	√
	WHEN AND WHY	√	√			
	DEPENDENCE	√	√			
	QUITTING ATTEMPTS	√	√			
	PLANS TO QUIT	√	√			
	REASONS TO QUIT	√	√			
	HAND-ROLLED	√	√			
	BRAND CHOICE	√	√			
	LAST PURCHASE	√	√			
	LIGHT/MILD	√	√			√
	PSYCHOSOCIAL BELIEFS	√	√			√
	HEALTH WARNINGS	√	√			√
	SHOW PACK SECTION	√	√			
	PERCEIVED RISK	√	√			√
	MINI-QUIT	√	√	√	√	√
	REASONS DID QUIT	√	√	√	√	√
	STAY QUIT					√
	LIGHT/MILD (NON-USERS)	√	√	√	√	√
	PSYCHOSOCIAL BELIEFS (NON-USERS)	√	√	√	√	√
	HEALTH WARNINGS (NON-USERS)	√	√	√	√	√√
BIDI	PAST & PRESENT FREQUENCY	√	√	√	√	
	DEPENDENCE	√	√			
	QUITTING ATTEMPTS	√	√			
	PLANS TO QUIT	√	√			
	REASONS TO QUIT	√	√			
	PSYCHOSOCIAL BELIEFS	√	√			√

Survey Sections		Survey Types				
		M	T	L	N	Q
	PERCEIVED RISK	√	√			√
	MINI-QUIT	√	√	√	√	√
	PSYCHOSOCIAL BELIEFS (NON-USERS)	√	√	√	√	√
SMOKED TOBACCO	REASONS TO QUIT	√	√			
	QUIT REPORTED					√
	KNOWLEDGE	√	√			√
	PSYCHOSOCIAL BELIEFS	√	√			√
	KNOWLEDGE (NON-USERS)			√	√	√
	PSYCHOSOCIAL BELIEFS (NON-USERS)			√	√	√
OTHER SMOKED TOBACCO	PAST & PRESENT FREQUENCY	√	√	√	√	√
ENVIRONMENTAL TOBACCO SMOKE		√	√	√	√	√
GENERAL CESSATION ASSISTANCE		√	√	√	√	√
ANTI-TOBACCO CAMPAIGNS		√	√	√	√	√
TOBACCO PROMOTION		√	√	√	√	√
TOBACCO INDUSTRY		√	√	√	√	√
ELECTRONIC CIGARETTES		√	√	√	√	√
MODERATORS		√	√	√	√	√
DEMOGRAPHICS		√	√	√	√	√
SURVEY CLOSING		√	√	√	√	√

4 Data Collection, Entry, Delivery, and Cleaning

4.1 Interviewer Training Workshop

The ITC Waterloo Project Manager (PM) prepared a training manual and schedule for the ITC Kenya training workshop. The facilitators for the interviewer training workshop included the ITC Waterloo Project Chief Principal Investigator (Prof. Geoffrey T. Fong), ITC Waterloo Managing Director and Senior Research Scientist (Dr. Anne Quah), and the ITC Research Scientist/Project Manager (Dr. Susan Kaai). The in-country Co-Principal Investigators, Dr. Lawrence Ikamari (University of Nairobi, PSRI) and Dr. Jane Nabongo (Kenya Medical Research Institute) also facilitated some of the training sessions. The training was conducted between March 11 and 16, 2018 at the University of Nairobi. Four Regional Coordinators, 19 Field Supervisors and 45 Interviewers were trained.

4.2 Data Collection

The period of the data collection was from April 09, 2018 to June 19, 2018. Data collection was slowed down due to the heavy rains and floods the country was experiencing at that time (see Appendix H). Other challenges which the fieldwork teams faced included:

- The two EAs (Ilandi Village EA-042701030100619 and Thange Village EA-04 27 03 02 02 016 1 9) that were bordering Chyulu National Park. These EAs are sparsely populated and bushy---as one interviewer put it "it was challenging to walk through the rocky terrain and the bushy thicket."
- The Kibwezi team faced the following challenges:
    - Heavy rains made the poor roads impassable

- Being a rainy season, most households who were involved in farming economic activities in their own farms or those who were casual laborers went to look for work outside their homes, hence it was difficult and challenging to find respondents at home.

4.3 Survey Process

In Wave 2, the ITC Survey protocol comprised nine main steps where an interviewer:

1. Approached a household and introduced him or herself
2. Completed the Household Recontact Form (HRF) or Household Enumeration Form (HEF).
3. Administered the household (H) survey to the head or key Informant.
4. Identified respondents for the individual interviews.
5. Obtained informed consent.
6. Administered the survey screeners in order to identify the appropriate survey to use.
7. Completed the individual surveys determined by the screening result.
8. Completed the individual outcome codes in the HRF/HEF Form.
9. Ended each interview and provided the token of appreciation.

4.4 Languages, Translation, and Length of Survey

The ITC Project pays a lot of attention to ensure that the translation into the local language(s) is of high quality. It must be accurate and have the closest possible equivalence to the English version. Proper translation plus work to ensure the accuracy and the quality of the translation of each of the English version could take 4-6 weeks. A standard ITC Project Translation Protocol is provided to the country translator.

At Wave 1, several different surveys were created depending on the user types, i.e., Smoked Survey (T), Smokeless Survey (L), Mixed-User Survey (M) and the Non-User Survey (N). The Household Survey (H) is specifically for the head of the household. All surveys were adapted by the Waterloo and Kenyan team to ensure that they were relevant to the Kenyan context (in English) and then translated to Kiswahili by a professional Kenyan translator and verified by a second Kenyan translation consultant.

At Wave 2, we used the same surveys that were developed at Wave 1 with a few modifications and new questions. Specifically, we added a few questions to evaluate the impact of the recently implemented pictorial health warnings. We also added a new survey, i.e., the Quitters Survey (Q) for Wave 1 tobacco user respondents who had quit at Wave 2. Most of the questions in this survey are similar to the other questionnaires with a few additions. The bilingual surveys (English and Kiswahili) were administered in the Kiswahili (national language) or English (official language) for those respondents who preferred to answer the survey in English. The surveys took approximately 60 minutes for tobacco users and approximately 30 minutes for non-users to complete.

4.5 Data Checking

At the end of each day, interviewers completed a self-check on the Household Surveys, Household Recontact Forms (HRFs), Household Enumeration Forms (HEFs), consent forms, screeners, and surveys they collected during the day. They reviewed all materials to determine whether anything was missed or skipped. The field

supervisor collected all completed fieldwork documents from the interviewer teams under his/her supervision and reviewed them before sending them via courier services to the Project Manager for further quality checks and filing.

4.6 Data Entry

Following data collection activity, the ITC Kenya data were manually entered into computer files using the freely available EpiData software (http://www.epidata.dk/). Data entry templates were programmed by our ITC programmer at the University of Waterloo and reviewed by the in-country Kenyan data manager. Data entry templates were programmed to ensure correct skip patterns were followed and to prevent data entry clerks from entering invalid values. Double data entry was performed by two separate data entry clerks. Each clerk entered the data once. Such duplicate data entry helped minimize data entry errors since it is unlikely that two different people will make the same data entry error for a given value.

4.7 Data Delivery and Cleaning

Once duplicate data entry was completed by the in-country data clerks, the data files were transferred securely to the University of Waterloo using ITC's secure internal website called SENDIT, which can only be accessed by users who have an account on the website. As an extra precaution to maintain security, data files were encrypted prior to uploading them to the website. Once the data were successfully transferred, the University of Waterloo data analyst commenced data cleaning. The data analyst conducted duplicate entry comparisons of the data files using SAS statistical software and identified discrepancies between the two data files. A list of these discrepancies was sent to the in-country data manager for verification and correction. The in-country data manager sent the corrections to the University of Waterloo data analyst for verification.

After discrepancies had been identified and corrections sent by the in-country data manager, the University of Waterloo data analyst conducted additional checks on the data to ensure that all skip patterns had been correctly followed and to ensure that the data did not contain invalid values. Respondent identifier codes were also checked thoroughly to ensure the data could be correctly linked within a survey wave and between waves over time. Any additional discrepancies that were identified were also sent back to the in-country data manager for verification. This back and forth communication between the University of Waterloo data analyst and the in-country data manager went on until the data were deemed clean by the University of Waterloo data analyst. Following data processing and cleaning, sampling weights were constructed for the dataset and the final, cleaned datasets were released to the country team, by posting them on the secure, internal ITC website.

4.8 Collecting Empty Tobacco Packs

During the data collection exercise, empty tobacco packs provided voluntarily by respondents who smoked cigarettes or bidis were collected by the interviewers and were subsequently handed to the Field Supervisors for storage.

4.9 Remuneration

For Wave 2, tobacco user and non-user respondents received soap or airtime* to the amount of KES 250 (3 USD) as a token of appreciation for their participation. The amount of token was set after country PI consulted with Kenyan government officials and stakeholders.

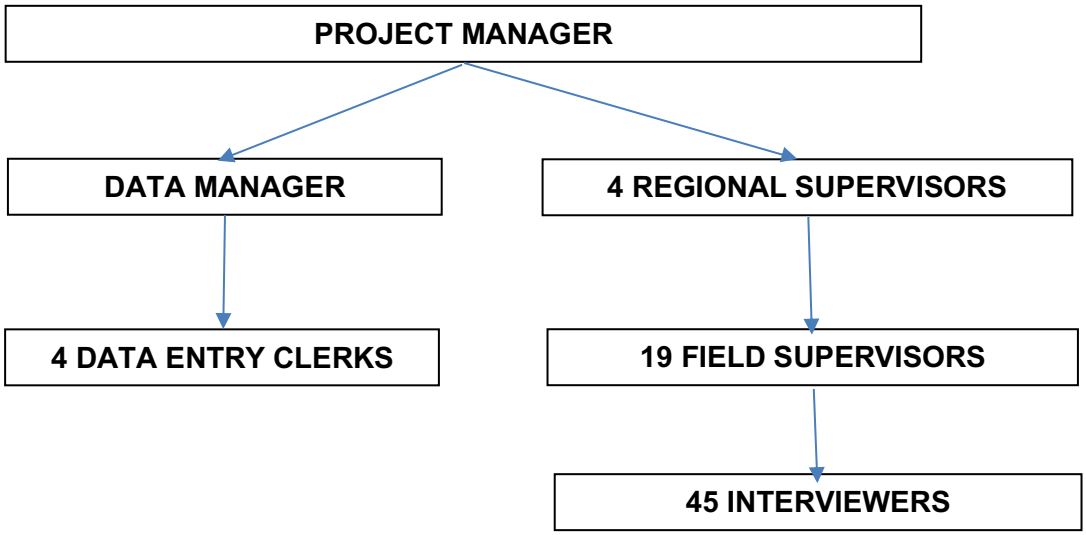
*All prepaid mobile phones in Kenya need 'talktime', 'credit', or 'airtime' so the user can make calls, send texts and if they have a smartphone, use data. This airtime must be bought before the phone is used and when the user runs out of airtime, they cannot use the phone until they have topped up with more airtime. Prepaid mobile phone service is the most popular and accessible mode of phone service in Kenya.

5 Monitoring and Quality Control

5.1 Management of Fieldwork Teams

The project fieldwork team consisted of four levels of management as shown below.

Figure 2: Management of ITC Kenya Project



Their responsibilities included the following activities:

Project Manager:

- Overseeing all aspects of the survey fieldwork & data quality
- Completing administrative responsibilities of the project
- Providing necessary guidance to team members
- Communicating with the University of Waterloo (UW) on every aspect of the project
- Overseeing data entry process and check data entry errors
- Checking and reporting fieldwork progress including sending the weekly update form to UW

Data Manager:

- Testing that the epi-data template sent from UW was correctly used
- Collecting and checking all completed forms and surveys
- Training and supervising the data entry clerks
- Overseeing the double data entry process
- Overseeing the checking, cleaning and compilation of data
- Transferring data from paper to computer (double entry), data cleaning and encrypting data set before sending it to the Data Management Centre (DMC) at UW

Regional Supervisors:

- Overseeing all aspects of the survey fieldwork to ensure that the fieldwork is executed as planned in the region
- Providing necessary guidance to team members in the respective regions and communicating any challenges to the Program Manager
- Monitoring the field Supervisors to ensure that they checked and signed the completed surveys
- Compiling field reports and submitting them to the Project Manager

Field Supervisors:

- Informing relevant local authorities about the survey
- Identifying HHs selected from each Village/EA and ensuring all interviewers are ready with the HH list for each Village/EA
- Managing surveys and consent forms, and ensuring that interviewers had all the materials for fieldwork
- Traveling with the team and addressing any questions or concerns from the interviewers
- Checking the quality of information given by respondents and recording by the interviewers
- Communicating with the Project Manager about progress and difficulties encountered in the field
- Completing the fieldwork progress report and sending it to the Project Manager
- Handling travel arrangements and other field logistics

Interviewers:

- Obtaining consent from each survey respondent
- Interviewing each respondent using the correct survey based on the smoking status of the respondent (from the Screener)
- Reporting any problems or concerns to the field supervisor

Each Province was covered by one team comprising 1 supervisor and 2-4 interviewers.

5.2 Interviewer Aids

Several interviewing aids were used to facilitate the administration of surveys. They include the following:

- **Manikin Flashcards:** There were two questions in each of the surveys that required the aid of a flashcard when the response options were read out, to save time and to facilitate interviewing. The flashcards included pictures of little man-like figures (manikins) with bipolar degrees of emotion (from negative to positive emotions). Researchers studying emotional responses found the use of these manikins to be helpful to respondents in rating their own emotional responses. (See Manikin Flashcards in Appendix D).
- **Reference Sheets:** Interviewers were also provided with reference sheets that included pictures of different tobacco products in the Kenyan market (Appendix E) and step-by-step guidelines/interview steps on how to conduct the data collection activity (see Flow Chart, Appendix C).

5.3 Monitoring & Quality Assurance

To ensure the accuracy and the quality of the ITC Kenya Survey, the fieldwork was monitored in several ways.

- **Regional and Field Supervisors**: The Field Supervisor travelled with each interviewer team and provided regular feedback to the interviewers. The Field Supervisor ensured that the survey protocol and data collection standards were being closely followed. Field Supervisors were available to address any questions or concerns from the interviewers. The Regional Supervisors monitored the quality of the fieldwork specifically to see that the Field supervisors executed their responsibilities well.
- **Identification Numbers:** Field Supervisors were instructed to ensure that households and respondents identification numbers were being properly filled and labels were correctly used.
- **Checking for Completeness:** At the end of each day, interviewers were required to perform a self-check on the surveys they completed that day. The interviewer reviewed the whole survey to determine whether any questions had erroneously been missed or skipped. The Field Supervisor then collected all completed surveys from the interviewers under their supervision and conducted a thorough check to ensure that they were correctly filled. If there were gaps, the Fields Supervisor asked the interviewer to re-visit the household member and complete the survey. Upon the completion of each survey in each village/ward the Field Supervisor gave the completed and checked surveys to Regional Supervisors who conducted a final check before sending them to the in-country Project Manager (PM), who then gave it to the data entry Manager.
- **Weekly Meetings**: The objectives of the meetings were to check and monitor the fieldwork progress, deal with matters arising from the field, and monitor the budget and the expenditures of the fieldwork and the data sending process. Issues related to exhausted EAs for replenishment, heavy rains, floods, where resistance was met, or any other challenges encountered in the field, were discussed in these meetings.
- **Progress Reports:** The in-country Project Manager provided regular email updates (to Waterloo PM), a weekly progress reports on quotas completed and a narrative report regarding the progress of the data collection exercise. He also informed the Waterloo PM about any concerns or problems that arose in the field.

5.4 Data Quality Control

The double entry of data proceeded in parallel with the data collection activity. In order to ensure the quality of the data collection process, the team used a multistage monitoring system:

- There were random visits by the in-country regional coordinators to monitor the interviewers in the field
- Recontact and replenishment households were randomly called to verify the information that the enumerators filled on the forms
- Field Supervisors cross-checked all completed enumeration and surveys daily to ensure that they were properly filled
- After data entry was completed, the Data Manager ran routine checks on the data sets and informed the in-country PM about any potential problems. When any issues arose the in-country PM contacted the supervisors using mobile phones. The Field Supervisors would then relay the message to the interviewers.

5.5 Handling Special Situations

- **Private Interviews:** The standard was that the participants would be privately interviewed alone; however in some circumstances some household members insisted on being present during the interview. If such a situation arose the interviewer would only proceed with the interview if the respondent was agreeable to having the person present during the interview.
- **A Proxy Interview:** A proxy interview is an interview conducted with another knowledgeable member of the household on behalf of the selected respondent. An example would be a woman answering the survey for her husband. Proxy interviews were not allowed in the ITC Kenya Survey.
- **Respondent was Unavailable:** If a respondent was unavailable, an appointment time was rescheduled to a time that was convenient to the respondent. Only four such attempts were allowed.
- **Substitution:** A substitution from the same household was allowed ONLY if a selected individual from the Non-User category (N status) had Individual Outcome Code 2 (Language barrier) or Code 3 (Health/mentally incapable) or Code 8 (Away from the household for the entire survey period).

6 Disposition Codes

6.1 Disposition Codes for Household Outcome Codes

1. Could not find
2. Vacant
3. Not a household (e.g., business premise)
4. Threat to safety
5. No contact- weather condition
6. No answer- 2 attempts
7. No answer- survey period ends
8. Household refusal
9. Language barrier
10. No one capable of answering
11. Enumeration prevented for other reasons
12. Enumerated

6.2 Disposition Codes for Individual Outcome Codes

1. Missed (after 4 attempts)
2. Language barrier
3. Health/Mentally incapable
4. Proxy refusal
5. Refusal
6. Incomplete (start, break-off)
7. Completed
8. Away for the entire survey period

6.3 Outcome Rates for ITC Kenya Wave 1 Survey

Number of households enumerated = 1876
Number of households enumerated with replenishment respondents = 941
Number of tobacco users interviewed (including quitters) = 1308
Number of non-users interviewed = 523
Total number of interviews= 1831

6.4 Sample Size and Representation Tables

A total of 1700 respondents were left in the data file after cleaning (i.e., removal of duplicates and out of frame households). These respondents were from 1594 households. Table 4 below shows a breakdown of the sample by smoking status and gender. Table 4 shows the total respondents participating in ITC Kenya Wave 2. Table 5 shows the Wave 1 to Wave 2 retention rates and initial tobacco use at the time of recruitment by Province.

Table 4: Total Sample of Respondents by Smoking Status and Gender

Sample Area (provinces)	Cigarette only		Mixed		Smokeless only		Quitter		Non-user	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
01	94	15	7	0	3	1	5	1	17	20
02	125	2	0	0	1	3	21	1	28	30
03	78	2	3	0	11	7	8	3	25	18
04	79	1	3	0	18	30	9	1	26	23
05	30	0	1	0	2	0	2	0	9	6
06	118	12	2	0	10	11	7	2	32	40
07	151	13	0	0	71	51	9	2	60	66
08	101	11	5	0	47	6	19	6	36	41
Subtotal A	**776**	**56**	**21**	**0**	**163**	**109**	**80**	**16**	**233**	**244**
Subtotal B	**832**		**21**		**272**		**96**		**477**	
Grand Total	**1698 (2 respondents were missing gender)**									

Table 5: Total Respondents Participating in ITC Kenya Wave 2 Survey

Sample	Tobacco Status	Cohort		
		Cohort 1 Recontact	Cohort 2 Replenishment	Total
National	Cigarette only	372	461	833
	Mixed	5	16	21
	Smokeless only	151	121	272
	Quitter	96	0	96
	Non-user	248	230	478

Table 6: Waves 1 to 2 Retention Rates by Initial Tobacco Use at the Time of Recruitment

Provinces	Tobacco User				Non-user				All Respondents			
	Lost		Retained		Lost		Retained		Lost		Retained	
	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
01 Nairobi	131	81.88	29	18.12	56	90.32	6	9.68	187	84.23	35	15.77
02 Central	56	35.90	100	64.10	28	44.44	35	55.56	84	38.36	135	61.64
03 Coast	79	69.91	34	30.09	36	78.26	10	21.74	115	72.33	44	27.67
04 Eastern	144	62.88	85	37.12	66	68.75	30	31.25	210	64.62	115	35.38
05 North Eastern	31	81.58	7	18.42	10	66.67	5	33.33	41	77.36	12	22.64
06 Nyanza	98	54.44	82	45.56	41	54.67	34	45.33	139	54.51	116	45.49
07 Rift Valley	195	55.24	158	44.76	62	45.93	73	54.07	257	52.66	231	47.34
08 Western	69	34.85	129	65.15	24	30.38	55	69.62	93	33.57	184	66.43
Overall	**803**	**56.27**	**624**	**43.73**	**323**	**56.57**	**248**	**43.43**	**1126**	**56.36**	**872**	**43.64**

Appendix A: ITC Kenya Wave 1 Sampling Design

The sampling frame for Wave 1 of the ITC Kenya Survey was the 2009 Kenya Population & Housing Census (KPHC) conducted by the Kenya National Bureau of Statistics (KNBS). According to that frame, the population of Kenya was first divided into 8 provinces; see Figure A1. Each province was then divided into districts (or wilaya) (see Figure A2), with a grand total of 158 districts. Each district was then further divided into divisions (or taarafa), which in turn were divided into locations (or mtaa) and then sub-locations (or mtaa mdogo). Finally, each of the over 7,000 sub-locations was divided into enumeration areas (EAs). These EAs consist on average of about 100 households, but their sizes varied quite a lot.

Figure A-1: Map of Kenya showing the 8 Provinces in 2009

Source:

https://www.google.com/search?q=map+of+kenya&oq=Map+of+kenya&aqs=chrome.0.0l6.2671j0j7&sourceid=chrome&ie=UTF-8

Figure A-2: Districts of Kenya in 2009

Source:

https://www.google.com/search?q=map+of+kenya&oq=Map+of+kenya&aqs=chrome.0.0l6.2671j0j7&sourceid=chrome&ie=UTF-8

Our original sampling design consisted in stratifying the population, and then sampling districts (stage I), locations (stage II), sub-locations (stage III), EAs (stage IV), households (stage V), and finally individual respondents (stage VI). However, several locations contained few sub-locations; yielding selection probabilities close to 1 at stage III. For other locations, stage III selection probabilities would have been much smaller, and such a scenario would have resulted in sampling weights that would have been highly variable; thus decreasing

precision. To avoid this, locations that contained too few sub-locations were pooled together or pooled with larger locations. This pooling yielded what we have called super-locations. In other words, super-locations are an artificial level between divisions and locations that we created. The same issue arose with sub-locations containing too few EAs, and a few sub-locations were thus pooled together (Table A1).

Table A-1: Overview of the Stratified 6–Stage Sampling Plan

Hierarchy	Explanation
Strata	8 Provinces yielded 12 Strata
Stage 1	Sample 1-5 districts within the 12 strata (n=21)
Stage 2	Sampled 1-2 super-locations within each of 21 districts (n=37)
Stage 3	Sampled 2 sub-locations within each of 36 super-locations, while 3 sub-locations were sampled via PPS in the remaining super-location (n=75)
Stage 4	Sampled 2-3 EAs per sub-location (n=148)
Stage 5	Every household within each sampled EA was assigned a unique ID number. Those household ID numbers were then randomly ordered within EA, and households were visited by interviewers following the random ordering
Stage 6	Sampled up to 4 tobacco users and 1 non-user per enumerated household. This was repeated until the quota of the EA was met i.e., 10 tobacco users and 4 non-users of tobacco

There were special considerations that were made in the sampling design. They are described below:

1. Because the North Eastern Province borders Somalia, there are various safety concerns. Additionally, there are other difficulties regarding interviewing respondents in this province. In the light of this, instead of randomly sampling from the 4 districts in that province, it was decided to sample the district of Garissa with probability one (i.e., purposive sampling); see Patton (1990) for more information on purposive sampling. Note that Garissa is one of the two largest districts of that province.
2. The districts of Migori, Kuria West and Kuria East in the province of Nyanza are Kenya's main tobacco-growing regions. It was thus decided to sample one of those three districts. To this end, the province of Nyanza was divided into two strata: the first stratum consisted of the districts of Migori, Kuria West and Kuria East, whereas the second stratum consisted of the 18 other districts of Nyanza; see Table A2).
3. The districts of Mombasa, Eldoret West, Eldoret East, Wareng, Busia, Teso North and Teso South are of particular interest because of potential tobacco smuggling. It was thus decided to sample 3 of those 4 districts, one per province. The district of Mombasa is the only such district in the Coast province, and it was thus sampled with probability one. As was done for the province of Nyanza, the provinces Western and Rift Valley were each divided into two strata; see Table A2.

Table A-2: Strata of the ITC Kenya Wave 1 Survey

Stratum #	Abbreviation	Province	Districts	Number of districts sampled in stratum *h*
1	CE	Central	All	2
2	CO1	Coast	All except Mombasa	1
3	CO2	Coast	Mombasa	1
4	EA	Eastern	All	3
5	NA	Nairobi	All	2
6	NE	North Eastern	Garissa	1
7	NY1	Nyanza	All except those in NY2	2
8	NY2	Nyanza	Migori, Kuria West, Kuria East	1
9	RV1	Rift Valley	All except those in RV2	4
10	RV2	Rift Valley	Eldort West, Eldoret East, Wareng	1
11	WE1	Western	All except those in WE2	2
12	WE2	Western	Busia, Teso North, Teso South	1

Allocation of EAs

Two to three EAs were randomly sampled with probability proportional to population size in each of the 75 sub-locations that were selected at stage three. Refer to Table A3.

Table A-3: Sample Design for ITC Kenya Wave 1 Survey

Province	Sampled Districts	Sampled EAs/Clusters (n)
Central	Murang'a South	8
Central	Murang'a North	8
Coast	Kilindini	8
Coast	Mombasa	4
Eastern	Kibwezi	8
Eastern	Machakos	8
Eastern	Meru South	8
Nairobi	Nairobi East	8
Nairobi	Nairobi North	8
Nyanza	Kisii Central	8
Nyanza	Migori	4
Nyanza	Nyando	8
Rift Valley	Baringo	8
Rift Valley	Eldoret West	4
Rift Valley	Kericho	8
Rift Valley	Naivasha	8
Rift Valley	Narok North	8
Western	Bungoma South	8
Western	Lugari	8
Western	Teso	4
North Eastern	Garissa	4
TOTAL	**21 Districts**	**148 EAs**

Sampling of Households and Enumeration

The interviewer teams would go to the selected clusters also known as enumeration areas (EAs). If a list of the dwelling units of an EA was not already available, they would make a new list following written guidelines from ITC (see details in Section 3.9). They would visit the dwelling units on the list in random order, and at each dwelling unit, once contact was made, enumerate the household (i.e., enumeration included listing all the household members and their age, sex, tobacco use, and relationship to head of household), and select individuals for interview, if any were eligible. They would conduct the interviews then or at a later visit. They would stop enumerating households in the EA when the target numbers of interviews for the EA were reached. Enumeration is important for estimating prevalence, which could be done with small bias even if not all of the expected 10,500 households were enumerated, as long as households were approached in random order and the enumeration data were carefully recorded even when there was no one eligible in the households.

Sampling of Individuals within a Household

In any enumerated household all tobacco users up to a maximum of 4 (randomly selected if necessary) were to be interviewed. If there were more than 4 users in a household, the interviewers sampled all female users and selected male users at random until 4 users were sampled. This procedure was meant to increase the potential number of female users in the sample. Smoking is much less prevalent among females as compared to males. Sampling was to continue until the cluster quota of users (10 users) was reached. In every 4th (Central and Eastern Provinces), 5th (Nairobi, Coast, Rift Valley, and North-Eastern Provinces), 7th (Nyanza and western Provinces) enumerated household, up to one randomly selected non–user (male or female) was to be interviewed until the quota of non-users (4 non-users) for the cluster was met. If a 4th (or 5th or 7th, as appropriate) household did not contain any users, interviewers were still expected to sample a non-user. Among non-users, an alternation procedure was designed to sample an equal number of males and females.

Appendix B: Random Table

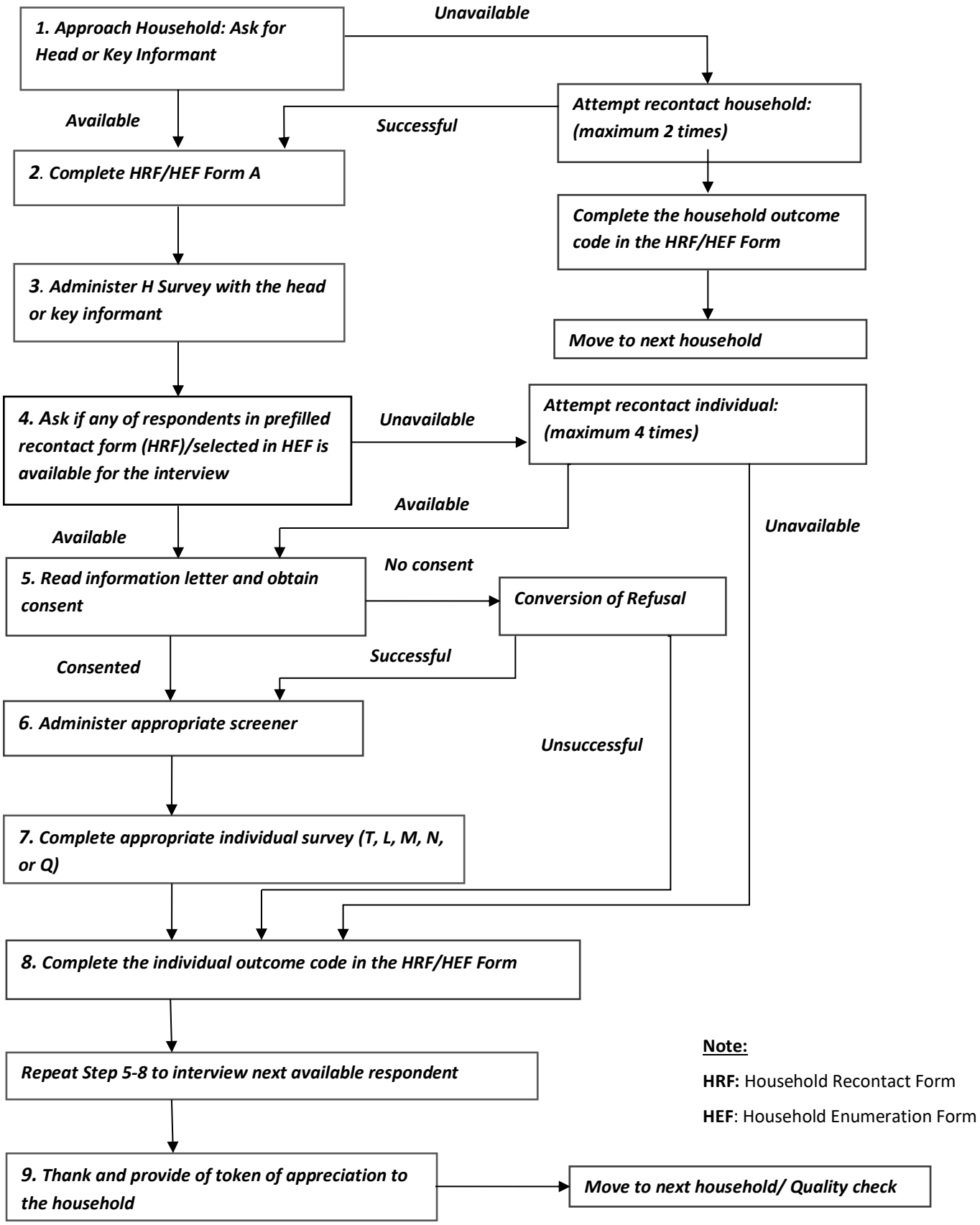
Table B-1: Random Order Sequence Table for Household Selection (A 1-page sample)

76 - 100 HH units	101-125 HH units	126-150 HH units	151-175 HH units	176-200 HH units	201-225 HH units	226-250 HH units	251-275 HH units	276-300 HH units	301-325 HH units	326-350 HH units	351-375 HH units	376-400 HH units	401-425 HH units	426-450 HH units	451-475 HH units	476-500 HH units	Male/ Female / No Non-User (M/F/ NNU)
																484	
					216	216	216	216	216	216	216	216	216	216	216	216	
															466	466	
64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	
				185	185	185	185	185	185	185	185	185	185	185	185	185	
												399	399	399	399	399	
					220	220	220	220	220	220	220	220	220	220	220	220	
											374	374	374	374	374	374	
39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	
							254	254	254	254	254	254	254	254	254	254	
						238	238	238	238	238	238	238	238	238	238	238	
														439	439	439	
24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	
							257	257	257	257	257	257	257	257	257	257	
															468	468	
								280	280	280	280	280	280	280	280	280	
									307	307	307	307	307	307	307	307	

Table B-2: Random Order Sequence Table for Household Selection (A 1-page sample)

501-525 HH units	526-550 HH units	551-575 HH units	576-600 HH units	Male/ Female/ No Non- User (M/F/ NNU)
338	43	425	86	
182	259	509	305	
104	481	72	210	
99	144	538	111	
429	128	26	530	
497	133	206	393	
384	489	290	394	
281	50	267	172	
116	288	142	361	
510	252	531	511	
266	548	411	279	
478	141	194	493	
480	395	264	342	
88	251	481	45	
177	379	101	271	
156	315	64	541	
133	471	124	591	
6	503	535	96	
506	535	259	231	
295	383	316	383	
471	373	3	519	

Appendix C: Flowchart for ITC Kenya Wave 2 Survey Interview Steps



Note:

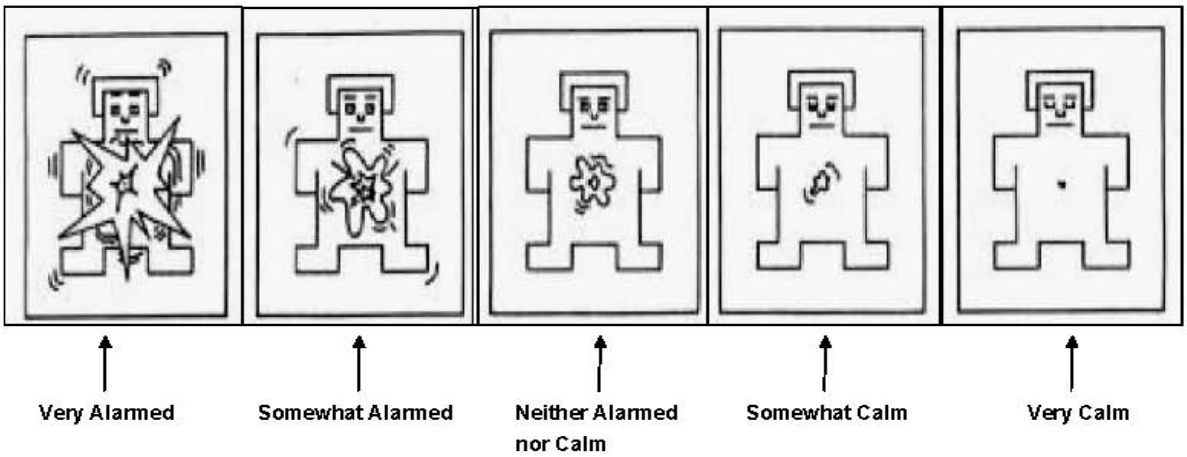
HRF: Household Recontact Form

HEF: Household Enumeration Form

Appendix D: Manikin Flashcards

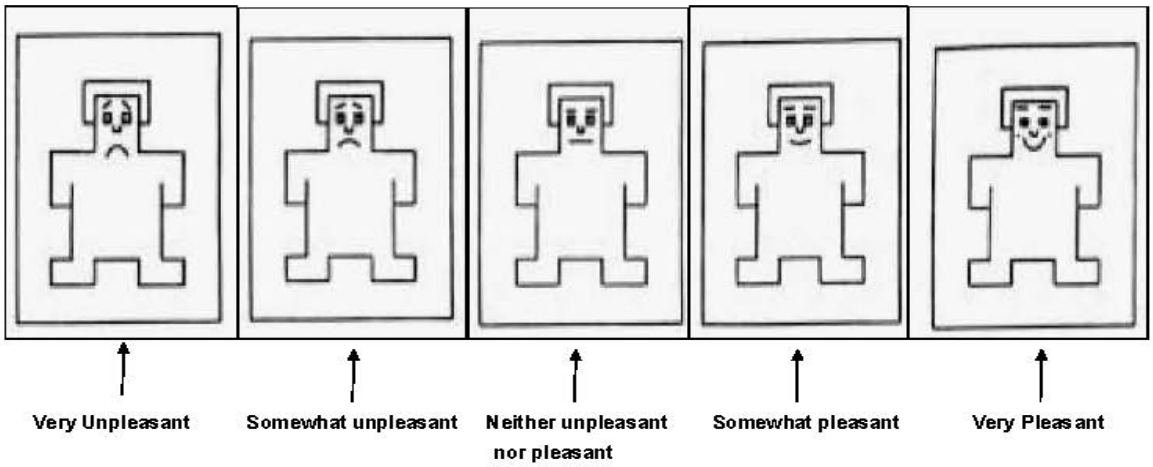
MANIKIN FLASHCARD 1

AROUSAL SCALE



MANIKIN FLASHCARD 2

VALENCE SCALE



Appendix E: Tobacco Products

Bidi – tobacco wrapped in tendu or temburni leaves (native to India)

:Source: https://www.google.com/search?biw=1200&bih=855&tbm=isch&sa=1&ei=Vf-XW9bLEsLujwTMyYLYBA&q=Bidi&oq=Bidi&gs_l=img.3..0j0i67k1j0l8.51029.51947.0.5214 8.4.4.0.0.0.0.80.261.4.4.0....0...1c.1.64.img..0.4.261....0.0AtOIbNWHrc#imgrc=OVY0RI6vVo_UMM

Kretek – a cigarette made of tobacco and cloves (native to Indonesia)

Source: https://www.google.com/search?biw=1200&bih=855&tbm=isch&sa=1&ei=if-XW_CRPMqkjwTWxbyACg&q=Kretek&oq=Kretek&gs_l=img.3..0l10.32059.33376.0.3 3745.6.6.0.0.0.0.115.412.5j1.6.0....0...1c.1.64.img..0.6.408...0i67k1j0i10k1.0.G5zBW0 dgErA#imgrc=7q2GEJyc-CQNBM

Cheroot – a cigar with both ends clipped and left open

Chutta – a coarsely made cheroot, typically homemade

Source: https://www.bing.com/images/search?view=detailV2&ccid=LgfddahG&id=6FB459B0226 31A6E678498200FAD0269E80E1842&thid=OIP.LgfddahG_GSmPJwNTbUp4gHaHa&mediaurl=http%3a%2f%2fcigarplace.biz%2fmedia%2fcatalog%2fproduct%2fcache%2f1%2fthumbnail%2f9df78eab33525d08d6e5fb8d27136e95%2fl%2fa%2fla_flor_dominicana_n.a.s._cheroot_single_cigar_ft%2fLa-Flor-Dominicana-N.A.S.-Cheroot-cigarplace.biz-

Hookah / water pipe – a pipe used to smoke shisha in which the smoke is cooled before inhalation by being drawn though water

Source:https://www.google.com/search?biw=1200&bih=855&tbm=isch&sa=1&ei=K_mXW5b4NIrgjwSSq57gCg&q=hookah&oq=hookah&gs_l=img.3..0j0i67k1l2j0l2j0i67k1j0j0i67k1j0l2.1784.5009.0.5217.8.6.1.1.1.0.69.367.6.6.0....0...1c.1.64.img..0.8.373...0i10k1.0.8S_oClTURvg#imgrc=kjdZFM7wHYwZX

Pipe – a reusable device used for smoking tobacco, consisting of a bowl, stem, and mouthpiece

Hookli – clay pipe (common in India)

Source: http://www.lotustalk.com/forums/f3/ca-smog-bottom-line-82913/

Cigar – a tightly rolled bundle of tobacco

Cigarillo – a short, narrow cigar

Sources: http://www.photo-dictionary.com/phrase/530/cigar.html#b, http://www.mikescigars.com/brands/chevere-cigarillo

<u>Electronic cigarettes / Vapes</u> – a battery powered device that delivers nicotine, flavor, and other chemicals (heat e-liquid)  Source: https://static1.squarespace.com/static/546c7443e4b0f54deccc32f3/t/57fe06926b8f5b365be198c1/1476265765268/Images+courtesy+of+Anna+Phillips	**<u>Heat-not-Burn (HNB) Products</u>** – a battery powered device that heats tobacco sticks to generate nicotine  Source: https://i2.wp.com/www.tobaccoreporter.com/wp-content/uploads/2017/11/iqos.jpg?resize=350%2C200&ssl=1

Appendix F: ITC Kenya Wave 2 Screeners

ITC KENYA WAVE 2 SURVEY– SCREENER 1: REPLENISHMENT **1**

Interviewer ID: ☐☐

ID: ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐☐ ☐ ☐ ☐☐☐☐☐ ☐☐☐
Province District Division Location Sub-Location EA No. EA Type EA Status Structure No. Household No.

Individual ID: ☐☐

Date of Survey: _____ (dd)/_____(mm)/_____(yy)

Start Time:_____am/pm **End Time:** _____am/pm

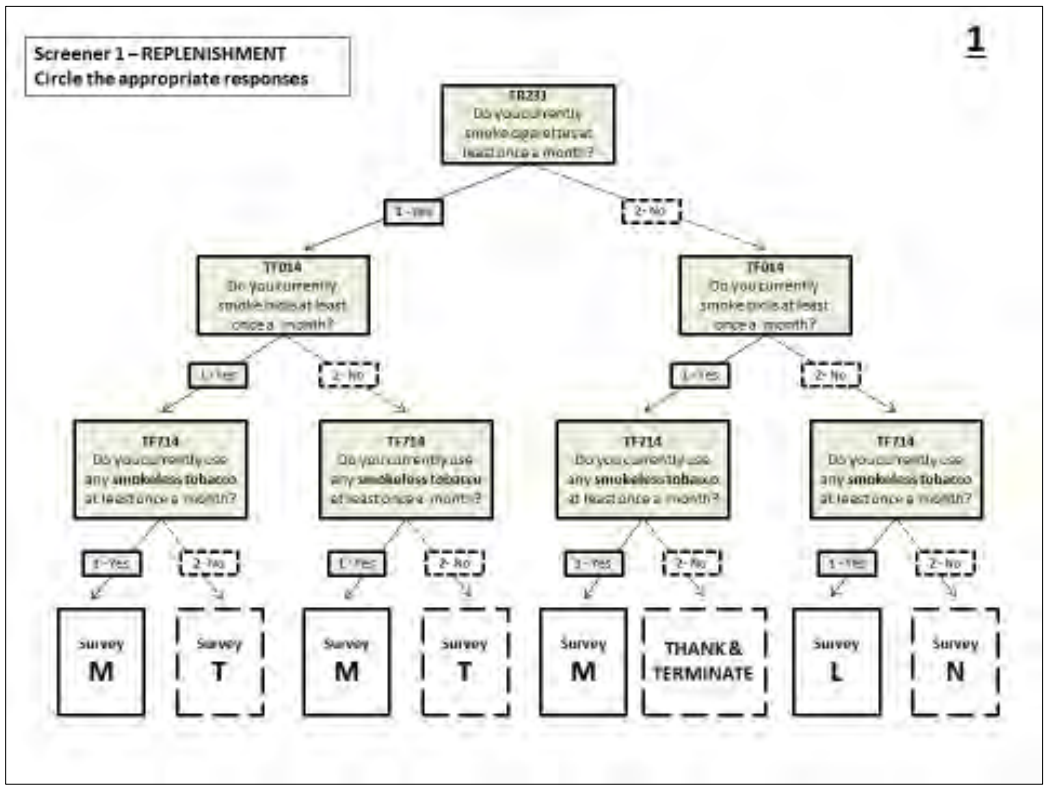
Please tick the correct box for each question below

1. Do you currently smoke cigarettes at least once a month (FR231)? Yes☐ No☐
   Je, kwa sasa unavuta sigara angalau mara moja kwa mwezi?
2. Do you currently smoke bidis at least once a month (TF014)? Yes☐ No☐
   Je, kwa sasa unavuta bidis angalau mara moja kwa mwezi?
3. Do you currently use any smokeless tobacco at least once a month?
   Includes nasal snuff, oral snuff, chewing tobacco, Kuber and other products (TF714)? Yes☐ No☐
   Je, kwa sasa unatumia tumbaku isiyovutwa angalau mara moja kwa mwezi?
   Bidhaa hizi zinahusisha ugoro wa puani, ugoro wa mdomoni, tumbaku ya kutafuna, Kuber au aina nyingine ya bidhaa zisizovutwa za tumbaku.

Use flowchart to select survey and then record here (either M, T, L or N): _____

Checked by:_____ (Name of Supervisor)

Date Checked:_____



ITC KENYA WAVE 2 SURVEY– SCREENER 2: W1 Tobacco User

2

Interviewer ID: ☐☐

INSERT THE ID CODE LABEL

Province	District	Division	Location	Sub-Location	E.A. No.	EA type	EA status	Structure No	Household No.
☐☐	☐☐	☐☐	☐☐	☐☐	☐☐☐	☐	☐	☐☐☐☐☐	☐☐☐

Individual ID: ☐☐

Date of Survey: _____ (dd)/_____(mm)/_____(yy)
Start Time:_____am/pm **End Time:** _____am/pm

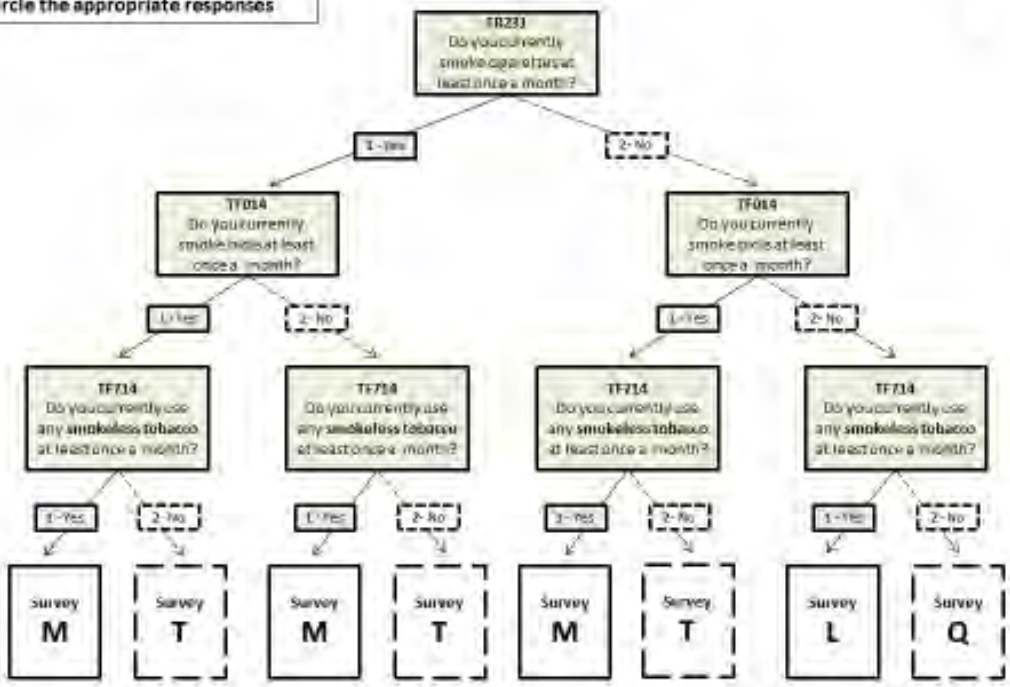
Please tick the correct box for each question below

1. Do you currently smoke cigarettes at least once a month (FR231)? Yes ☐ No ☐
   Je, kwa sasa unavuta sigara angalau mara moja kwa mwezi?
2. Do you currently smoke bidis at least once a month (TF014)? Yes ☐ No ☐
   Je, kwa sasa unavuta bidis angalau mara moja kwa mwezi?
3. Do you currently use any smokeless tobacco at least once a month?
   Includes nasal snuff, oral snuff, chewing tobacco, Kuber and other products (TF714)? Yes ☐ No ☐
   Je, kwa sasa unatumia tumbaku isiyovutwa angalau mara moja kwa mwezi?
   Bidhaa hizi zinahusisha ugoro wa puani, ugoro wa mdomoni, tumbaku ya kutafuna, Kuber au aina nyingine ya bidhaa zisizovutwa za tumbaku.

Use flowchart to select survey and then record here (either M, T, L or Q): _____
Checked by:_____ (Name of Supervisor)
Date Checked:_____

Screener 2 – W1 Tobacco User
Circle the appropriate responses

2



ITC KENYA WAVE 2 SURVEY– SCREENER 3: W1 Non-User of Tobacco

3

Interviewer ID: ☐☐

INSERT THE ID CODE LABEL

Province	District	Division	Location	Sub-Location	E.A. No.	EA type	EA status	Structure No	Household No.
☐☐	☐☐	☐☐	☐☐	☐☐	☐☐☐	☐	☐	☐☐☐☐☐	☐☐☐

Individual ID: ☐☐

Date of Survey: _____ (dd)/_____(mm)/_____(yy)

Start Time._____am/pm End Time: _____am/pm

Please tick the correct box for each question below

1. Do you currently smoke cigarettes at least once a month (FR231)? Yes ☐ No ☐
   Je, kwa sasa unavuta sigara angalau mara moja kwa mwezi?
2. Do you currently smoke bidis at least once a month (TF014)? Yes ☐ No ☐
   Je, kwa sasa unavuta bidis angalau mara moja kwa mwezi?
3. Do you currently use any smokeless tobacco at least once a month?
   Includes nasal snuff, oral snuff, chewing tobacco, Kuber and other products (TF714)? Yes ☐ No ☐
   Je, kwa sasa unatumia tumbaku isiyovutwa angalau mara moja kwa mwezi?
   Bidhaa hizi zinahusisha ugoro wa puani, ugoro wa mdomoni, tumbaku ya kutafuna, Kuber au aina nyingine ya bidhaa zisizovutwa za tumbaku.

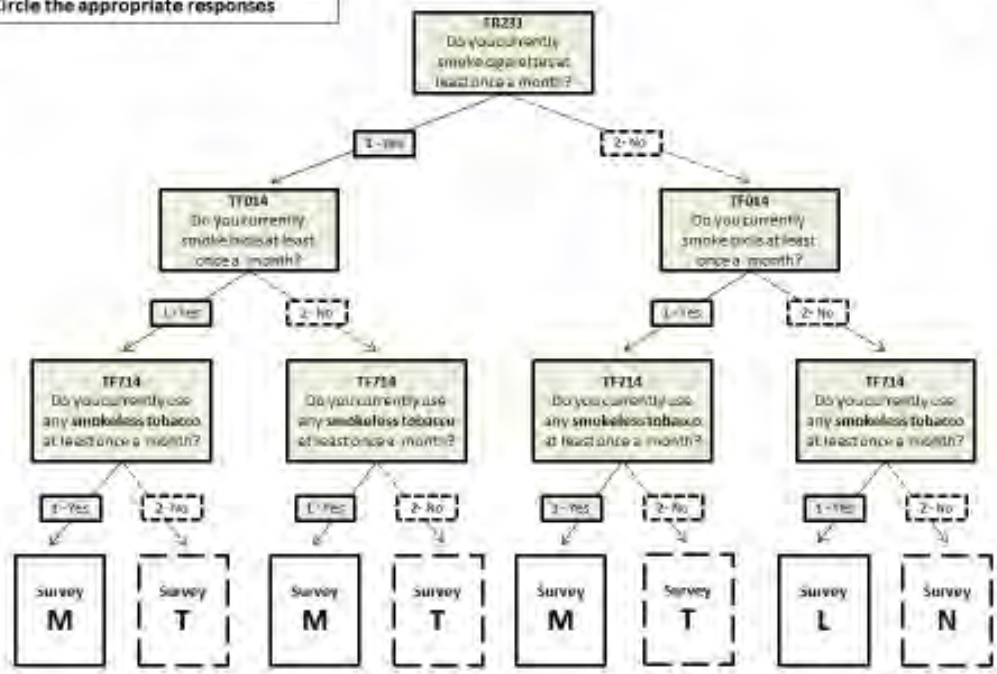
Use flowchart to select survey and then record here (either M, T, L or N): _____

Checked by:_____ (Name of Supervisor)

Date Checked:_____

3

Screener 1 – W1 Non-User of Tobacco
Circle the appropriate responses

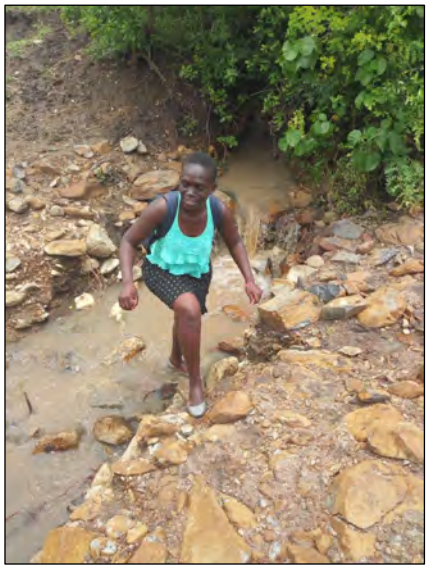


Appendix G: Example of a Prefilled Recontact Label for Wave 2

NAI: [DANDORA PHASE IV 'B'2 Village/Estate]
01 02 01 01 02 **037 2 9** 00187 086
Individual ID: **03** ITC KE2: Screener **3**
Wave 1 cohort (Last tobacco status -- non-smoker)

OLD code: 01020101020372900186100 01

Appendix H: Photos of Interviewers in the Field



Appendix I: Weight Construction (Waves 1 and 2)

The chapter on weights construction is provided below.

Sampling Design and Weight Construction for the International Tobacco Control (ITC) Kenya Survey

C. Boudreau[1,2], Y. Li[2,3], Y. Fan[2,3] and M. Thompson[1,2]

This technical report details the sampling design and weight construction for Waves 1 and 2 of the International Tobacco Control (ITC) Kenya Survey. The ITC Kenya Survey was designed as a prospective longitudinal survey of a national representative random sample of approximately 1500 smokers and 600 non-smokers.

This technical report is organized as follows: section 1 describes the sampling design of the ITC Kenya Survey, section 2 details the construction of the sampling weights for the data from Waves 1 and 2, and section 3 contains Wave 1 smoking prevalence estimates and a description of how they were computed.

1 Sampling design

The ITC Kenya Survey was designed as a prospective longitudinal study, and its sampling design at Wave 1 was chosen to yield representative random samples of tobacco users and non-users 15 years of age and older. Respondents were interviewed in October–December 2012. All interviews were conducted face-to-face.

To qualify for the study, respondents were required to be 15 years old or more. Those that currently smoked cigarettes or bidis, or used any form of smokeless tobacco at least once a month were considered to be tobacco users. Those that were randomly selected completed the Smoked Tobacco Survey, the Smokeless Tobacco User Survey or Mixed User Survey depending on their tobacco consumption habit. All other qualifying respondents were considered to be non-users, and those that were randomly selected completed the Tobacco Non-User Survey.

1.1 Special considerations

1. Because it is neighbouring Somalia, there are various safety concerns and other difficulties in interviewing respondents in the province of North Eastern. Instead of randomly sampling from the 4 districts in that province, it was decided to sample the district of Garissa with probability one (i.e., purposive sampling); see Patton (1990) for more information on purposive sampling. Note that Garissa is one of the two largest districts of that province.

[1]Dept. of Statistics & Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada.
[2]Data Management Core (DMC) – ITC Project, University of Waterloo.
[3]Dept. of Psychology, University of Waterloo.
[4]This document was created using LaTeX, and last updated on Oct 8, 2020

2. The districts of Migori, Kuria West and Kuria East in the province of Nyanza are Kenya's main tobacco-growing regions. It was thus decided to sample one of those three districts. To this end, the province of Nyanza was divided into two strata: stratum NY2 consisted of the districts of Migori, Kuria West and Kuria East, whereas stratum NY1 consisted of the other 18 districts of Nyanza; see table 1.

3. The districts of Mombasa, Eldoret West, Eldoret East, Wareng, Busia, Teso North and Teso South are of particular interest because of potential tobacco smuggling. It was thus decided to sample 3 of those 4 districts, one per province. The district of Mombasa is the only such district in the Coast province, and it was thus sampled with probability one. As was done for the province of Nyanza, the provinces Western and Rift Valley were each divided into two strata; see table 1.

Stratum #	Abbr.	Province	Districts	n_h^I
1	CE	Central	all	2
2	CO1	Coast	all except Mombasa	1
3	CO2	Coast	Mombasa	1
4	EA	Eastern	all	3
5	NA	Nairobi	all	2
6	NE	North Eastern	Garissa	1
7	NY1	Nyanza	all except those in NY2	2
8	NY2	Nyanza	Migori, Kuria West, Kuria East	1
9	RV1	Rift Valley	all except those in RV2	4
10	RV2	Rift Valley	Eldoret West, Eldoret East, Wareng	1
11	WE1	Western	all except those in WE2	2
12	WE2	Western	Busia, Teso North, Teso South	1

n_h^I = number of districts sampled in stratum h

Table 1: Strata of the ITC Kenya Survey.

1.2 Sampling frame

The sampling frame for Wave 1 of the ITC Kenya Survey is the 2009 Kenya Population & Housing Census (KPHC) conducted by the Kenya National Bureau of Statistics (KNBS). According to that frame, the population of Kenya was first divided into 8 provinces; see figure 1. Each province was then divided into districts (or *wilaya*), with a grand total of 158 districts. Each district was then further divided into divisions (or *taarafa*), which are in turn divided into locations (or *mtaa*) and then sub-locations (or *mtaa mdogo*). Finally, each of the over 7,000 sub-locations was divided into enumeration areas (EAs). These EAs consist on average of about 100 households, but this varies quite a lot.

Our original sampling design consisted in stratifying the population, and then sampling districts (stage I), locations (stage II), sub-locations (stage III), EAs (stage IV), households (stage V) and finally individual respondents (stage VI). However, several locations contain few sub-locations; yielding selection probabilities close to 1 at stage III. For other locations, stage III selection probabilities would have been much

smaller, and such a scenario would have resulted in sampling weights that would have been highly variable; thus decreasing precision. To avoid this, locations that contained too few sub-locations were pooled together or pooled with larger locations. This pooling yielded what we have called super-locations. In other words, super-locations are an artificial level between divisions and locations that we have created. The same issue arose with sub-locations containing too few EAs, and a few sub-locations were thus pooled together.

Figure 1: Provinces of Kenya.

1.3 Sampling of units at Wave 1

To accommodate the special considerations listed in section 1.1, Kenya was first divided into 12 strata; see table 1. The Wave 1 sample of 1500 tobacco users and 600 non-users was divided equally between 150 EAs/clusters; hence, 10 tobacco users and 4 non-users were to be randomly sampled from each EA. The sampling of units for Wave 1 of ITC Kenya then proceeded as follows:

Stage I: Within stratum h ($h = 1, \ldots, 12$), n_h^I districts were randomly sampled with probability proportional to population size (PPS); see table 1. This stratified PPS sampling without

Figure 2: Districts of Kenya.

replacement was done using the Hanurav-Vijayan algorithm; see Hanurav (1967) and Vijayan (1968) for details.

Stage II: Within selected district d ($d = 1, \ldots, 21$), n_d^{II} super-locations were randomly selected using PPS sampling without replacement. In the majority of districts, 2 super-locations where sampled. For a handful of districts however, a single super-location was selected. Hence, a grand total of 37 super-locations were sampled.

Stage III: 2 sub-locations were then randomly sampled with PPS in 36 of the 37 super-locations selected at stage II; whereas, 3 sub-locations were sampled via PPS in the remaining super-location (this exception was the result of the sub-locations being small in that particular super-location). Since a handful of super-locations contained few sub-locations, a PPS method (Murthy, 1957)was used at this stage.

Stage IV: Within a selected sub-location ℓ ($\ell = 1, \ldots, 75$), n_ℓ^{IV} EAs were randomly sampled with PPS. For the majority of sub-locations $n_\ell^{IV} = 2$, but there were a few instances where the EAs of a sub-location were small and thus 3 were selected instead.

Stage V: Every household within each sampled EA was assigned a unique ID number. Those ID numbers were then randomly ordered within the EA, and households were visited by interviewers following the random ordering.

Stage VI: Finally, up to 4 tobacco users and 1 non-user per household were randomly selected and interviewed. This was repeated until both quotas were met for that EA; see ITC Kenya Survey Fieldwork Manual for details on stages V and VI. A non-user could only be selected for interview in every κ-th household visited, where κ was equal to 4 in the Central and Eastern Provinces; equal to 5 in Nairobi, Coast, Rift Valley and North-Eastern Provinces; and equal to 7 in the Nyanza and Western Provinces. Non-users were chosen so as to alternate in gender: if a male non-user was recruited in the 1st dwelling unit enumerated, a female non-user would be recruited in the $\kappa + 1$-st dwelling unit enumerated, etc.

A total of 5454 households were enumerated at Wave 1 of the ITC Kenya Survey. Individual interviews were conducted in 1776 of those, yielding a sample of 1427 tobacco users and of 571 non-users; for a total of 1998 individual respondents.

1.4 Sampling of units at Wave 2

Strong efforts were made to recontact and interview respondents from Wave 1. Where this was not possible, those lost to follow-up were replaced through a replenishment sample, conducted where possible within the same cluster or EA. Additional dwelling units within the same EA were visited, and their households enumerated. Up to 4 tobacco users and 1 non-user per newly enumerated household were randomly selected and interviewed. This was repeated until both quotas were met for that EA. It was intended that a non-user could be selected for interview only in every κ-th household visited, where κ was equal to 4 in the Central and Eastern Provinces; equal to 5 in Nairobi, Coast, Rift Valley and North-Eastern Provinces; and equal to 7 in the Nyanza and Western Provinces. Non-users were ro be chosen so as to alternate in gender.

Where insufficient households were available in the EA for enumeration and sampling, a new EA was added in the same sub-location to complete the replenishment process.

2 Weight construction

2.1 General comments about weight construction

As with most survey weights, the ITC Kenya weights are constructed to correct and adjust for sample mis-representation caused by unequal sampling probabilities, frame error (i.e., under-coverage and multiplicity), and non-response as well as improving precision of estimates through the use of auxiliary information (e.g., tobacco usage prevalences). We briefly describe these key concepts of weight construction in this section, but refer the reader to Levy & Lemeshow (2008), chapter 16, for more detailed information.

At their base, sampling weights are defined as the inverse of inclusion probabilities, and thus adjust for sample misrepresentation caused by unequal sampling probabilities. For example, a tobacco user residing with three others has a probability of inclusion of one, while a tobacco user residing with four others has a probability of inclusion of 0.8.

Great efforts are made to create a complete/perfect sampling frame (i.e., a frame that would include all members of the target population, without duplicate and without any erroneous inclusions[1]). However, this is rarely achieved and, consequently, some members of the target population are not part of the sampling frame (i.e., have a 0 probability of being selected). This is referred to as frame under-coverage, and can result in non-coverage bias. To reduce non-coverage bias in the ITC Kenya Survey, post-stratification adjustments were performed on the sampling weights to ensure that, for each sex/age/province group for male tobacco users and sex/age group for female tobacco users, the totals of the sampling weights equal estimated number of such tobacco users based on the enumeration data; see step 3 in section 2.2.2. A similar adjustment was used for respondents that do not use any tobacco products; see step 4 in section 2.2.3.

If non-respondents behave differently than respondents, inference based solely on the sample of respondents will be biased unless adjustments are made. The greater the expected proportion of non-response, the greater this bias is likely to be. In the ITC Kenya Survey, the post-stratification adjustments described in the above paragraph also adjust for non-coverage bias. It should be noted that if data are missing completely at random (MCAR, see Little & Rubin, 2002) within each sex/age group, then non-response bias will be completely eliminated. Realistically, non-response bias is greatly reduced, but not eliminated in the ITC Kenya Survey.

All weights for the ITC Kenya Survey were computed using the statistical software SAS (http://www.sas.com).

2.2 Wave 1 weights

2.2.1 Household weights

Enumeration household weights (EHWT) were computed for the 5454 households enumerated (i.e., contacted and listed) at Wave 1. Computation of the EHWT weights proceeded as follows:

Step 1: Each enumerated household was first assigned an enumeration area (EA) level weight $w_j^{(1)}$. This weight corresponds to the inverse of the selection probability of the jth household given that its EA was sampled. Formally,

$$w_j^{(1)} = \frac{M_{k(j)}}{m_{k(j)}}$$

where j stands for the jth household, $k(j)$ denotes the EA to which household j belongs, $M_{k(j)}$ is the total number of households in that EA (obtained from the 2009 census), and $m_{k(j)}$ is the number of households enumerated in that same EA. Hence, in any given EA, all households will have the same $w_j^{(1)}$ weight.

[1] Erroneous inclusions refers to units that are not part of the target population, but included in the sampling frame.

Note: For one of the 105 EAs, $m_{k(j)} > M_{k(j)}$, and the above formula could not be used as it would have resulted in $w_j^{(1)} < 1$. The computation of the $w_j^{(2)}$ weights for the households of that of EA is described in step 2 below.

Step 2: The $w_j^{(1)}$ weights were then multiplied by a factor to obtain the $w_j^{(2)}$ weights. This factor corresponds to the inverse of the selection probability of the kth EA. This yielded the $w_j^{(2)}$ weights, which are formally given by

$$\begin{aligned} w_j^{(2)} &= \underbrace{\frac{N_h^I}{n_h^I N_d^{II}}}_{\substack{\text{selecting}\\ \text{district } d}} \times \underbrace{\frac{N_d^{II}}{n_d^{II} N_e^{III}}}_{\substack{\text{selecting}\\ \text{super-location } e}} \times \underbrace{\frac{N_e^{III}}{n_e^{III} N_\ell^{IV}}}_{\substack{\text{selecting}\\ \text{sub-location } \ell}} \times \underbrace{\frac{N_\ell^{IV}}{n_\ell^{IV} N_k^{V}}}_{\substack{\text{selecting}\\ \text{EA } k}} \times w_j^{(1)} \\ &= \frac{N_h^I}{n_h^I\, n_d^{II}\, n_e^{III}\, n_\ell^{IV}\, N_k^V} \times w_j^{(1)} \end{aligned}$$

where

$$\begin{aligned} n_h^I &= \#\text{ of districts sampled in stratum } h \\ n_d^{II} &= \#\text{ of super-locations sampled in district } d \\ n_e^{III} &= \#\text{ of sub-locations sampled in super-location } e \\ n_\ell^{IV} &= \#\text{ of EAs sampled in sub-location } \ell \\ N_h^I &= \text{population of stratum } h \\ N_d^{II} &= \text{population of district } d \\ N_e^{III} &= \text{population of super-location } e \\ N_\ell^{IV} &= \text{population of sub-location } \ell \\ N_k^V &= \text{population of EA } k \end{aligned}$$

Notes:

- N_h^I, N_d^{II}, N_e^{III}, N_ℓ^{IV} and N_k^V were all obtained from the 2009 Kenya Population and Housing Census.
- $N_h^I/(n_h^I N_d^{II})$ is the inverse of the selection probability of district d, with population N_d^{II}, given that n_h^I districts were sampled in stratum h.
- $N_d^{II}/(n_d^{II} N_e^{III})$ is the inverse of the selection probability of super-location e, with population N_e^{III}, given that n_d^{II} super-locations were sampled in district d.
- $N_e^{III}/(n_e^{III} N_\ell^{IV})$ is the inverse of the selection probability of sub-location ℓ, with population N_ℓ^{IV}, given that n_e^{III} sub-locations were sampled in super-location e.
- $N_\ell^{IV}/(n_\ell^{IV} N_k^V)$ is the inverse of the selection probability of EA k, with population N_k^V, given that n_ℓ^{IV} EAs were sampled in sub-location ℓ.

For the EA with $m_{k(j)} > M_{k(j)}$ mentioned above, computation of the $w_j^{(2)}$ weights proceeded as follows:

Step A) Compute $r_k = M_k/(m_k N_k^V)$ for all other EAs in the same super-location

Step B) Compute the average of the r_k's from step A, and let w_k^* be that average

Step C) The $w_j^{(2)}$ weight for all households in EA k was then taken to be

$$w_j^{(2)} = \frac{N_h^I}{n_h^I \, n_d^{II} \, n_e^{III} \, n_\ell^{IV}} \times w_k^*$$

Step 3: Let H_j^{15} be the total number of enumerated individuals, aged 15 and older, in the jth household. Multiplying H_j^{15} by $w_j^{(2)}$ and summing over all enumerated households in stratum h yields:

$$\widehat{N}_h^{15} = \sum_{j \in U_h} w_j^{(2)} \, H_j^{15}$$

which is an estimator of the total population, aged 15 and older, of stratum h ($h = 1, \ldots, 12$). The final step in the computation of the EHWT weights thus consists in multiplying the $w_j^{(2)}$ weights by a factor to ensure that

$$\sum_{j \in U_h} w_j^{\text{EHWT}} \, H_j^{15} = N_h^{15}$$

where N_h^{15} is the total population, aged 15 and older, of stratum h obtained from the 2009 census; see table 2. Consequently,

$$w_j^{\text{EHWT}} = w_j^{(2)} N_h^{15} / \widehat{N}_h^{15}$$

Stratum	**Population** (N_h^{15})
NA	2,187,366
CE	2,805,639
CO2	352,605
EA	3,300,875
NE	103,426
NY2	172,634
RV2	523,494
WE2	302,837
CO1	1,564,694
NY1	2,770,572
RV1	4,952,439
WE1	1,988,495

Table 2: Population, aged 15 and older, by stratum.

Interviewed household weights (IHWT) were computed for the 1776 households (i.e., 32.6% of the households enumerated at Wave 1) where one or more (i.e., up to 4) respondents completed one of the individual surveys (i.e., the cigarette, smokeless, mixed tobacco or non-user survey). To this end, we first need to divide the households into 2 groups: user households and non-user households. A household

is said to be a non-user household if none of its occupants, aged 15 and older, use tobacco products; otherwise, it is classified as a user household. Computation of the IHWT weights proceeded as follows,

$$w_j^{\text{IHWT}} = \begin{cases} w_j^{\text{EHWT}} \times \dfrac{UH^*_{k(j)}}{UH_{k(j)}} & \text{if household } j \text{ is a tobacco user household} \\ w_j^{\text{EHWT}} \times \dfrac{NH^*_{k(j)}}{NH_{k(j)}} & \text{if household } j \text{ is a non-user household} \end{cases}$$

where w_j^{EHWT} is the EHWT weight of household j computed above, $UH_{k(j)}$ is the number of user households interviewed in the EA of household j, and $UH^*_{k(j)}$ is the number of user households enumerated in that same EA. Similarly, for non-user households, $NH_{h(j)}$ is the number of non-user households interviewed in the EA of household j, and $NH^*_{h(j)}$ is the number of non-user households enumerated in that same EA.

2.2.2 Individual weights for tobacco users

Weights were computed for the 1427 respondents that completed one of the following 3 individual surveys: Smoked Tobacco Survey, Smokeless Tobacco User Survey or Mixed User Survey. Computation of those weights proceeded as follows:

Step 1: Each tobacco user was first assigned a within-household weight $w_i^{(1)}$, which corresponds to the inverse of his/her selection probability amongst eligible tobacco users of the same sex in his/her household. Formally, the $w_i^{(1)}$ weights are given by

$$w_i^{(1)} = \begin{cases} \dfrac{\#\text{male users}_{j(i)}}{\#\text{male users intvw}_{j(i)}} & \text{if } i^{\text{th}} \text{ respondent is a male} \\ \dfrac{\#\text{female users}_{j(i)}}{\#\text{female users intvw}_{j(i)}} & \text{if } i^{\text{th}} \text{ respondent is a female} \end{cases}$$

where $j(i)$ denotes the household in which respondent i lives, $\#\text{male users}_{j(i)}$ is the number of eligible male tobacco users in that same household, $\#\text{male user intvws}_{j(i)}$ is the number of such household members that were interviewed, and similarly for female tobacco users. Note: $w_i^{(1)}$ was equal to 1 for most respondents, and can only be greater than 1 in cases where more than 4 eligible tobacco users resided in the same household.

Step 2: Each $w_i^{(1)}$ weight was then multiplied by the corresponding IHWT weight. The resulting weights are labelled $w_i^{(2)}$, and are formally defined as

$$w_i^{(2)} = w_i^{(1)} \times w_{j(i)}^{\text{IHWT}}$$

where $w_{j(i)}^{\text{IHWT}}$ is the IHWT weight of the jth household (computed in section 2.2.1).

Step 3: A post-stratification adjustment was then performed to calibrate the $w_i^{(2)}$ weights to estimated prevalences of tobacco usage by sex/age/province groups; see section 3 for a description of how those prevalences were estimated. To this end, age was divided into 4 intervals; i.e., [15, 25),

[25, 35), [35, 50) and [50, 100). Males were further subdivided according to the province they live in. This yielded the first 28 sex/age/province cells of table 3, where some provinces and/or age groups were collapsed because they contained too few respondents. Females were not subdivided by provinces, as the prevalence of tobacco usage is very low amongst women. This yielded the remaining 4 sex/age group cells of table 3.

For respondents in cell C_k, this post-stratification adjustment consisted in multiplying their $w_i^{(2)}$ weights by a factor $\widehat{N}_k^{\text{user}}/t_k$ to produce calibrated $w_i^{(3)}$ weights. These $w_i^{(3)}$ weights are such that their sum over all respondents in cell C_k is equal to $\widehat{N}_k^{\text{user}}$, the estimated number of tobacco users in that cell. Formally,

$$w_i^{(3)} = w_i^{(2)} \times \frac{\widehat{N}_k^{\text{user}}}{t_k} = w_i^{(2)} \times \frac{\widehat{N}_k^{\text{user}}}{\sum_{i \in C_k} w_i^{(2)}}$$

where $\widehat{N}_1^{\text{user}}, \ldots, \widehat{N}_{32}^{\text{user}}$ are given in column 6 of table 3, and C_k is the set of all respondents in cell k ($k = 1, \ldots, 32$).

Step 4: Finally, the $w_i^{(3)}$ weights were rescaled for use in pooled regression or logistic regression analyses where age, gender and tobacco use status are covariates, as well as to facilitate comparisons with other ITC countries. To this end, the 1427 respondents were divided into 2 groups: male tobacco users ($g = 1$), and female tobacco users ($g = 2$). The weights were then rescaled to have mean equal to 1 in each group. This yielded the $w_i^{(4)}$ weights, which are formally defined as

$$w_i^{(4)} = w_i^{(3)} \times \frac{n_g}{\sum_{i \in S_g} w_i^{(3)}} ,$$

where S_g is the set of all respondents in group g, and n_g is the size of that sample; i.e.,

$$n_g = \begin{cases} 1183 & \text{if } g = 1 \\ 244 & \text{if } g = 2 \end{cases}$$

Note: the coefficient of variation (cv) of the $w_i^{(4)}$ weights is 1.42 for male tobacco users ($g = 1$), and 0.78 for female tobacco users ($g = 2$).

2.2.3 Individual weights for non-users

Weights were computed for the the 571 respondents that do not use any tobacco product, and thus completed the individual Tobacco Non-User Survey. Computation of those weights proceeded similarly to that of the weights for tobacco users; i.e.:

Step 1: Each non-user was first assigned a within-household weight $w_i^{(1)}$, which corresponds to the inverse of his/her selection probability amongst eligible non-users in his/her household. Formally, the $w_i^{(1)}$ weights are given by

$$w_i^{(1)} = \frac{\#\text{non-user}_{j(i)}}{\#\text{non-user intvw}_{j(i)}}$$

where $j(i)$ denotes the household in which respondent i lives, #non-user$_{j(i)}$ is the number of household members aged 15 and older that do not use any tobacco products, and #non-user intvw$_{j(i)}$ is the number of such household members that were interviewed (i.e., completed the individual Tobacco Non-User Survey).

Notes:

- #non-user$_{j(i)}$ was capped at 3 to control the variability of the weights.
- According to the sampling plan, #non-user intvw$_{j(i)}$ should be 1 for all households. However, there were few instances where 2 non-users from the same household were interviewed.

Step 2: Each $w_i^{(1)}$ weight was then multiplied by the corresponding IHWT weight. The resulting weights are labelled $w_i^{(2)}$, and are formally defined as

$$w_i^{(2)} = w_i^{(1)} \times w_{j(i)}^{\text{IHWT}}$$

where $w_{j(i)}^{\text{IHWT}}$ is the IHWT weight of the jth household (computed in section 2.2.1).

Step 3: Since a non-user residing with one or more tobacco users could only be selected while the non-user quota was opened, the $w_i^{(2)}$ weights of such non-users were multiplied by the estimated probability that the non-user quota was opened when interviewing a random household in EA k. This yielded the $w_i^{(3)}$ weights, which are formally defined as

$$w_i^{(3)} = \begin{cases} w_i^{(2)} & \text{if } i^{\text{th}} \text{ respondent lives in an household with only non-users} \\ w_i^{(2)} \times \dfrac{NH^*_{k(j)}}{NH_{k(j)}} & \text{if } i^{\text{th}} \text{ respondent lives in an household with 1 or more tobacco users} \end{cases}$$

where $NH_{k(j)}$ and $NH^*_{k(j)}$ were defined in the computation of the IHWT weights, and are respectively the number of non-user households enumerated in the EA of household j and the number of non-user households with interviews in the EA of household j. Note that the design intended that the non-user quota be open for user or non-user households approximation one time in κ, where κ was equal to 4 in the Central and Eastern Provinces; equal to 5 in Nairobi, Coast, Rift Valley and North-Eastern Provinces; and equal to 7 in the Nyanza and Western Provinces; the formula above accounts for possible departures from that approximation by using the reciprocal of the non-user household sampling rate in the EA.

Step 4: A post-stratification adjustment was then performed to calibrate the $w_i^{(3)}$ weights to the estimated number of individuals who do not use any tobacco products by sex/age groups; see section 3 for a description of how those numbers were estimated. To this end, age was divided into 4 intervals; i.e., [15, 25), [25, 35), [35, 50), [50, 100). This yielded the 8 sex/age cells of table 4. Note that we did not further divide by province, as our sample of non-users is much smaller that our sample of tobacco users. Hence, further subdividing by province would have resulted in cells with too few respondents, which could have in turn resulted in less stable weights.

For respondents in cell C_k, this post-stratification adjustment consisted in multiplying their $w_i^{(3)}$ weights by a factor $\widehat{N}_k^{\text{non-user}}/t_k$ to produce calibrated $w_i^{(4)}$ weights. These $w_i^{(4)}$ weights are such

that their sum over all respondents in cell C_k is equal to $\widehat{N}_k^{\text{non-user}}$, the estimated number of individuals, aged 15 and older, that do not use any tobacco products in that cell. Formally,

$$w_i^{(4)} = w_i^{(3)} \times \frac{\widehat{N}_k^{\text{non-user}}}{t_k} = w_i^{(3)} \times \frac{\widehat{N}_k^{\text{non-user}}}{\sum_{i \in C_k} w_i^{(3)}}$$

where the $\widehat{N}_k^{\text{non-user}}$ are given in column 5 of table 4, and C_k is the set of all respondents in cell k ($k = 1, \ldots, 8$).

Step 5: Finally, the $w_i^{(4)}$ weights were rescaled for use in pooled regression or logistic regression analyses where age, gender and tobacco use status are covariates, as well as to facilitate comparisons with other ITC countries. To this end, the 571 respondents were divided into 2 groups: non-users living with 1 or more tobacco users ($g = 1$), and non-users living with only other non-users ($g = 2$). The weights were then rescaled to have mean equal to 1 in each group. This yielded the $w_i^{(5)}$ weights, which are formally defined as

$$w_i^{(5)} = w_i^{(4)} \times \frac{n_g}{\sum_{i \in S_g} w_i^{(4)}}$$

where S_g is the set of all respondents in group g, and n_g is the size of that sample; i.e.,

$$n_g = \begin{cases} 103 & \text{if } g = 1 \\ 468 & \text{if } g = 2 \end{cases}$$

Note: the coefficient of variation (cv) of the $w_i^{(4)}$ weights is 1.31 for non-users living with 1 or more tobacco users ($g = 1$), and 1.47 for non-users living with only other non-users ($g = 2$).

2.3 Wave 2 weights

2.3.1 Longitudinal household weights

The interviewed household weights (IHWT), computed in section 2.2.1, have to be adjusted for attrition between Waves 1 and 2. To this end, let j stand for the jth household and w_j^{IHWT} be the IHWT computed in section 2.2.1. The *Waves 1–2 longitudinal household weight* for the jth tobacco user household is then given by

$$w_j^{\text{L12}} = w_j^{\text{IHWT}} \times \sum_{l \in U_{k(j)}^1} w_l^{\text{IHWT}} \Big/ \sum_{l \in U_{k(j)}^2} w_l^{\text{IHWT}} \tag{1}$$

where $k(j)$ denotes the EA to which household j belongs, $U_{k(j)}^1$ is the set of tobacco user households in the kth EA for whom one or more household members were interviewed at Wave 1, and $U_{k(j)}^2$ is the subset of those households for whom one or more household members were interviewed at Wave 2. Note that the denominator in (1) is the sum of the w_j^{IHWT}'s, computed in section 2.2.1, over all tobacco user

households in the kth EA where one or more household member was retained and interviewed at Wave 2. Basically, (1) ensures that

$$\sum_{j \in U_k^2} w_j^{\text{L12}} = \sum_{j \in U_k^1} w_j^{\text{IHWT}}$$

for all EAs. In other words, (1) ensures that, for each EA, the sum of the IHWT is the same at Waves 1 and 2.

Computation of the *Waves 1–2 longitudinal household weights* for non-user households was handled the same way. Hence, w_j^{L12}, the Waves 1–2 longitudinal household weights for the jth non-user household, is given by

$$w_j^{\text{L12}} = w_j^{\text{IHWT}} \times \sum_{l \in N_{k(j)}^1} w_l^{\text{IHWT}} \Big/ \sum_{l \in N_{k(j)}^2} w_l^{\text{IHWT}} \tag{2}$$

where $N_{k(j)}^1$ is the set of non-user households in the kth EA for whom one or more household members were interviewed at Wave 1, and $N_{k(j)}^2$ is the subset of those households for whom one or more household members were interviewed at Wave 2. As for tobacco user households, the above ensures that, for each EA, the sum of the IHWT is the same at Waves 1 and 2.

2.3.2 Longitudinal weights for tobacco users

Starting with $w_i^{(0)}$, the within-household weight of the ith respondent (i.e., weight $w_i^{(1)}$ computed in step 1 of section 2.2.2), computation of the 624 *Waves 1–2 longitudinal weights for tobacco users* proceeded as follows:

Step 1: Each $w_i^{(0)}$ weight was first multiplied by the corresponding Waves 1–2 longitudinal household weights computed in the previous section. The resulting weights are labelled $w_i^{(1)}$, and are formally defined as

$$w_i^{(1)} = w_i^{(0)} \times w_{j(i)}^{\text{L12}}$$

where $j(i)$ denotes the household to which individual i belongs to, and $w_{j(i)}^{\text{L12}}$ is the Waves 1–2 longitudinal household weight of that j^{th} household (computed in section 2.3.1).

Step 2: A post-stratification adjustment was then performed to calibrate the $w_i^{(1)}$ weights to estimated prevalences of tobacco usage by sex/age/province groups; see section 3 for a description of how those prevalences were estimated. To this end, age was divided into 4 intervals; i.e., [15, 25), [25, 35), [35, 50) and [50, 100). Note that for longitudinal weights, the age at the time the respondent was recruited (i.e., Wave 1 in the present case) is used. Males were further subdivided according to the province they live in. This yielded the first 28 sex/age/province cells of table 3, where some provinces and/or age groups were collapsed because they contained too few respondents. Females were not subdivided by provinces, as the prevalence of tobacco usage is very low amongst women. This yielded the remaining 4 sex/age group cells of table 3.

For respondents in cell C_k, this post-stratification adjustment consisted in multiplying their $w_i^{(1)}$ weights by a factor $\widehat{N}_k^{\text{user}}/t_k$ to produce calibrated $w_i^{(2)}$ weights. These $w_i^{(2)}$ weights are such

that their sum over all respondents in cell C_k is equal to $\widehat{N}_k^{\text{user}}$, the estimated number of tobacco users, aged 15 and older, in that cell. Formally,

$$w_i^{(2)} = w_i^{(1)} \times \frac{\widehat{N}_k^{\text{user}}}{t_k} = w_i^{(1)} \times \frac{\widehat{N}_k^{\text{user}}}{\sum_{i \in C_k} w_i^{(1)}}$$

where $\widehat{N}_1^{\text{user}}, \ldots, \widehat{N}_{32}^{\text{user}}$ are given in column 6 of table 3, and C_k is the set of all respondents in cell k ($k = 1, \ldots, 32$).

Step 3: Finally, the $w_i^{(2)}$ weights were rescaled for use in pooled regression or logistic regression analyses where age, gender and tobacco use status are covariates, as well as to facilitate comparisons with other ITC countries. To this end, the 624 respondents were divided into 2 groups: male tobacco users ($g = 1$), and female tobacco users ($g = 2$). The weights were then rescaled to have mean equal to 1 in each group. This yielded the $w_i^{(3)}$ weights, which are formally defined as

$$w_i^{(3)} = w_i^{(2)} \times \frac{n_g}{\sum_{i \in S_g} w_i^{(2)}} ,$$

where S_g is the set of all respondents in group g, and n_g is the size of that sample; i.e.,

$$n_g = \begin{cases} 517 & \text{if } g = 1 \\ 107 & \text{if } g = 2 \end{cases}$$

2.3.3 Longitudinal weights for non-users

Starting with $w_i^{(0)}$, the within-household weight of the ith respondent (i.e., weight $w_i^{(1)}$ computed in step 1 of section 2.2.3), computation of the 248 *Waves 1–2 longitudinal weights for non-users* proceeded as follows:

Step 1: Each $w_i^{(0)}$ weight was first multiplied by the corresponding Waves 1–2 longitudinal household weights computed in section 2.3.1. The resulting weights are labelled $w_i^{(1)}$, and are formally defined as

$$w_i^{(1)} = w_i^{(0)} \times w_{j(i)}^{\text{L12}}$$

where $j(i)$ denotes the household to which individual i belongs to, and $w_{j(i)}^{\text{L12}}$ is the Waves 1–2 longitudinal household weight of that j^{th} household (computed in section 2.3.1).

Step 2: Since a non-user residing with one or more tobacco users could only be selected while the non-user quota was opened, the $w_i^{(1)}$ weights of such non-users were multiplied by the estimated probability that the non-user quota was opened when interviewing a random household in EA k at Wave 1. This yielded the $w_i^{(2)}$ weights, which are formally defined as

$$w_i^{(2)} = \begin{cases} w_i^{(1)} & \text{if } i^{\text{th}} \text{ respondent lives in an household with only non-users} \\ w_i^{(1)} \times \dfrac{NH^*_{k(j)}}{NH_{k(j)}} & \text{if } i^{\text{th}} \text{ respondent lives in an household with 1 or more tobacco users} \end{cases}$$

where $NH_{k(j)}$ and $NH^*_{k(j)}$ were defined in section 2.2.1

Step 3: A post-stratification adjustment was then performed to calibrate the $w_i^{(2)}$ weights to the estimated number of individuals who do not use any tobacco products by sex/age groups; see section 3 for a description of how those numbers were estimated. To this end, age was divided into 4 intervals; i.e., [15, 25), [25, 35), [35, 50), [50, 100). Note that for longitudinal weights, the age at the time the respondent was recruited (i.e., Wave 1 in the present case) is used. This yielded the 8 sex/age cells of table 4. Note that we did not further divide by province, as our sample of non-users is much smaller that our sample of tobacco users. Hence, further subdividing by province would have resulted in cells with too few respondents, which could have in turn resulted in less stable weights.

For respondents in cell C_k, this post-stratification adjustment consisted in multiplying their $w_i^{(2)}$ weights by a factor $\widehat{N}_k^{\text{non-user}}/t_k$ to produce calibrated $w_i^{(3)}$ weights. These $w_i^{(3)}$ weights are such that their sum over all respondents in cell C_k is equal to $\widehat{N}_k^{\text{user}}$, the estimated number of individuals, aged 15 and older, that do not use any tobacco products in that cell. Formally,

$$w_i^{(3)} = w_i^{(2)} \times \frac{\widehat{N}_k^{\text{non-user}}}{t_k} = w_i^{(2)} \times \frac{\widehat{N}_k^{\text{non-user}}}{\sum_{i \in C_k} w_i^{(2)}}$$

where $\widehat{N}_k^{\text{non-user}}, \ldots, \widehat{N}_k^{\text{non-user}}$ are given in column 5 of table 4, and C_k is the set of all respondents in cell k ($k = 1, \ldots, 8$).

Step 4: Finally, the $w_i^{(3)}$ weights were rescaled for use in pooled regression or logistic regression analyses where age, gender and tobacco use status are covariates, as well as to facilitate comparisons with other ITC countries. To this end, the 248 respondents were divided into 2 groups: non-users living with 1 or more tobacco users ($g = 1$), and non-users living with only other non-users ($g = 2$). The weights were then rescaled to have mean equal to 1 in each group. This yielded the $w_i^{(4)}$ weights, which are formally defined as

$$w_i^{(4)} = w_i^{(3)} \times \frac{n_g}{\sum_{i \in S_g} w_i^{(3)}}$$

where S_g is the set of all respondents in group g, and n_g is the size of that sample; i.e.,

$$n_g = \begin{cases} 47 & \text{if } g = 1 \\ 201 & \text{if } g = 2 \end{cases}$$

2.3.4 Wave 2 cross-sectional weights

Retention rates varied widely across the provinces, and in some cases within provinces. Replenishment was carried out to try to replenish dropouts within their same EAs, from newly enumerated households. However, for many of the sampled areas, enumeration data were not provided for households newly approached at the replenishment stage, if they were not interview households, thus making it not possible to compute replenishment cross-sectional weights analogous to the Wave 1 cross-sectional weights. Thus Wave 2 cross-sectional weights were computed for the combined recontact and replenishment samples.

This section can be further divided into four parts: a) cross sectional interview household weights b) cross-sectional inflation weights for tobacco users, c) cross-sectional inflation weights for non-users, and d) cross-sectional analytic weights.

a) **Wave 2 cross-sectional interview household weights:**

Recall that a household is said to be a non-user household if none of its occupants (aged 15 and older) use tobacco products; otherwise, it is classified as a user household. Note that it is possible for a household that was classified as non-user at Wave 1 to be classified as a user household at Wave 2 if one or more of its occupants is now using tobacco. There were no cases of households needing to be re-classified in the Wave 2 recontact data.

Since quitters are considered as users for weight calculation purposes in ITC Kenya as in many other ITC surveys, the reverse is not possible, and a household was classified a user household at Wave 1, would remain so at Wave 2.

Let w_j^{EHWT1} be the EHWT of the jth Wave 1 enumerated household computed in section 2.2.1. Computation of the Wave 2 cross-sectional interview household weight within most EAs proceeded as follows.

We computed the sum S_{eh}^{EHWT1} of w_j^{EHWT1} over all Wave 1 enumerated households j, in EA e crossed with household type h. Then within each EA e and household type h, we gave the jth Wave 2 interview household an interview household weight of w_j^{IHWT2} that equals S_{eh}^{EHWT1} divided by the number of Wave 2 interview households.

Thus, the Wave 2 interview households in the EA represent all the households originally represented by the EA's Wave 1 enumerated households.

Exceptions occurred in cases where there was a new EA introduced at Wave 2, or when an EA with interviews in Wave 1 had none in Wave 2. Then where possible we ensured that the sum of the w_j^{EHWT1} in the EA's sublocation crossed with household type is the same as the sum of the Wave 2 interview household weights in the sublocation crossed with household type.

For example, suppose we have a sublocation s and household type h where both exceptions occur: there are new EAs introduced at Wave 2, and some EAs with interview households at Wave 1 have none at Wave 2 crossed with household type h. Households of type h that were Wave 1 interview households in EAs used at Wave 1 and Wave 2 have a value w_j^{IHWT1} of the Wave 1 interview household weight, specific to their EA and household type. Within a new EA crossed with household type h, we assigned to each interview household a value of w_j^{IHWT1} equal to the average of w_j^{IHWT1} in the sublocation s among Wave 1 interview households of Type h in EAs present in both waves. Then we rescaled the values of w_j^{IHWT1} for all Wave 2 interview households of type h in the sublocation s to sum to S_{sh}^{EHWT1}, the sum of w_j^{EHWT1} over enumerated households of type h in the sublocation s at Wave 1, producing Wave 2 interview household weights w_j^{IHWT2}.

Then if S_{sh}^{IHWT2} is the sum of w_j^{IHWT2} over all interview households in Wave 2 in sublocation s crossed with household type h, it should be equal to S_{sh}^{EHWT1}.

There were 22 new EAs in Wave 2: 5 in Province 1 (Nairobi), 3 in Province 2 (Central), 1 in Province 3 (Coast), 2 in Province 4 (Eastern), 1 in Province 5 (North Eastern), 4 in Province 6 (Nyanza), 5 in Province 7 (Rift Valley) and 1 in Province 8 (Western).

There were 13 EAs used in Wave 1 that were not visited in Wave 2: 10 in Province 4(Eastern), corresponding to district 0421, and 3 in Province 6 (Nyanza), with two corresponding to sublocation 0614010203. In these cases we used the above procedure, matching respectively the sums of the w_j^{IHWT2} and the w_j^{EHWT1} in Province 4 crossed with household type and (because of high variability after matching at the location level) in district 0614 crossed with household type.

In Province 8 (Western), w_j^{IHWT1} values in EAs 001 and 002 in sublocation 0820010602 were anomalous, and user households in these EAs were treated like households in a new EA.

In Province 1 (Nairobi), in sublocation 0102010102, one EA value of w_j^{IHWT1} was anomalous, and households in this EA were treated like households in a new EA.

There were additional cases of high variability of w_j^{IHWT2} for certain household types in Provinces 3, 6 and 8. The following remedies were adopted:

- In Province 3 (Coast), we replaced the w_j^{IHWT2} values of the non-user interview households of the sublocation 0302010301 by the average of the w_j^{IHWT2} of the non-user interview households over the two EAs of that sublocation.
- In Province 6 (Nyanza), we replaced the w_j^{IHWT2} values of the user interview households of the sublocation 0606020601, by the average of the $w_j^{\text{IHWT2}}$of the user interview households over the two EAs of that sublocation.
- In Province 8 (Western), we replaced the w_j^{IHWT2} values of the non-user interview households of the sublocation 0820010601 by the average of the w_j^{IHWT2} of the non-user interview households over the two EAs of that sublocation; similarly for the user interview households.

The construction of individual weights proceeded separately for tobacco users and for non-users. Following the implementation of the procedures below, two individuals missing the sex variable were assigned the average weight of the individual's age group and user status, within EA crossed with household type.

b) **Wave 2 cross-sectional inflation weights for tobacco users:** Computation of these cross-sectional weights proceeded the same way as for the Wave 1 weights for tobacco users (computed in section 2.2.2):

   (i) The Wave 2 cross-sectional interview household weights (computed above) were used as a basis instead of the Wave 1 interview household weights of section 2.2.1.

   (ii) The within-household weight $w_i^{(1)}$ was calculated as in section 2.2.2 for replenishment tobacco users; the $w_i^{(1)}$ weight from Wave 1 was used for recontact tobacco-users.

   (iii) Each $w_i^{(1)}$ weight was then multiplied by the corresponding w_j^{IHWT2} weight. The resulting weights are labelled $w_i^{(2)}$.

   (iv) The $w_i^{(2)}$ weights were calibrated by sex/age groups using the updated calibration figures of Table 5 instead of the Wave 1 calibration figures. In addition, age at Wave 2 was used instead of age at recruitment.

   (v) The resulting inflation weights are variable bDE72915v in the data set.

c) **Wave 2 cross-sectional inflation weights for non-users:** Computation of these cross-sectional weights proceeded in a similar way as for the Wave 1 weights for non-users (computed in section 2.2.3)

   (i) The Wave 2 cross-sectional interview household weights (computed above) were used as a basis instead of the Wave 1 interview household weights of section 2.2.1.

   (ii) The within-household weight $w_i^{(1)}$ was calculated as in section 2.2.3 for replenishment non-users; the $w_i^{(1)}$ weight from Wave 1 was used for recontact non-users.

(iii) Each $w_i^{(1)}$ weight was then multiplied by the corresponding w_j^{IHWT2} weight. The resulting weights are labelled $w_i^{(2)}$.

(iv) Pre-calibration weights were calculated as follows. The pre-calibration $w_i^{(3)}$ weight for a non-user living in a non-user household was the $w_i^{(2)}$ weight. For a non-user living in a user household, the pre-calibration $w_i^{(3)}$ weight was $w_j^{(2)}$ times $X=$ (average w_j^{IHWT2} for non-user households in their province) /(average w_j^{IHWT2} for user-households in their stratum) to account for the fact that the rate of interviewing non-users in interview user households was roughly equal to $1/X$ times the rate of interviewing users in interview non-user households. This adjustment makes the $w_i^{(3)}$ weights of similar magnitude for non-users from user households and from non-user households.

(v) The pre-calibration weights were calibrated by sex/age groups using the updated calibration figures of Table 6 instead of the Wave 1 calibration figures. In addition, age at Wave 2 was used instead of age at recruitment.

(vi) The resulting inflation weights are variable bDE72915v in the data set.

d) **Wave 2 cross-sectional analytic weights:**

Finally, the cross-sectional inflation weights were rescaled to produce cross-sectional analytic weights, for use in pooled regression or logistic regression analyses where age, gender and tobacco use status are covariates, as well as to facilitate comparisons with other ITC countries. To this end, the tobacco users were divided into 2 groups: male tobacco users ($g = 1$), and female tobacco users ($g = 2$). The weights were then rescaled to have mean equal to 1 in each group. The non-users were divided into 2 groups: non-users living in user households ($g = 3$), and non-users living in non-user households ($g = 4$).

The resulting analytic weights are variable bDE72919v in the data set.

Note: the coefficient of variation (cv) of the analytic weights is 1.03 for male tobacco users ($g = 1$), 0.95 for female tobacco users ($g = 2$), 0.86 for non-users living in user households ($g = 3$) and 0.95 for non-users living in non-user households($g = 4$).

3 Prevalence estimates

This section details how estimates for the prevalences of tobacco usage, per gender/age/province group, and the number of tobacco users were computed. For the time of Wave 1, these estimates are given in table 3, and were used to calibrate the sampling weights of tobacco users (see step 3 of section 2.2.2). This section also describes how we estimated the numbers of individuals, per gender/age group, who do not use any tobacco product, and the corresponding prevalences. For the time of Wave 1, these estimates are given in table 4, and were used to calibrate the sampling weights of non-users (see step 4 of section 2.2.3). All estimates for the time of Wave 1 are based on the 5454 household enumeration forms completed at Wave 1 of the ITC Kenya Survey.

Estimation of population totals and averages from survey data is usually very straightforward. In the present case however, we had the additional difficulty that age was missing for 442 individuals, and that

tobacco use status was missing for another 27 individuals[2]. Failure to take this into account would have led to under-estimation of the number of tobacco users, and likely bias prevalence estimates.

In section 3.1, we describe how calculations would have proceeded if there were no missing values; whereas, in section 3.2, we describe the procedure we actually used to obtain the figures in tables 3 and 4, which take into accounts the missing status and ages.

3.1 Assuming no missing values

Let k be one of the sex/province(s)/age group combinations for males, or sex/age group combinations for females, of table 3. The estimated number of tobacco users for the kth combination ($k = 1, \ldots, 32$) is given by

$$\widetilde{N}_k^{\text{user}} = \sum_{j \in U_h} w_j^{\text{EHWT}} \, HU_j^k$$

where h denotes the province (or provinces since some were merged) corresponding to the kth combination, w_j^{EHWT} is the EHWT of the jth household (computed in section 2.2.1), and HU_j^k is the number of enumerated tobacco users in the household j that fall within the sex/age group corresponding to the kth combination. Note that $HU_j^k = 0$ if household j contains no such individuals. Lastly, the sum is over all enumerated households in the hth province(s) for males, and over all enumerated households nationwide for females (since all 8 provinces were pooled together). Likewise, the estimated total number of individuals for the kth combination is given by

$$\widetilde{N}_k^{\text{total}} = \sum_{j \in U_h} w_j^{\text{EHWT}} \, H_j^k$$

where H_j^k is the total number of enumerated individuals in the jth household that fall within the sex/age group corresponding to the kth combination. Combining $\widetilde{N}_k^{\text{user}}$ and $\widetilde{N}_k^{\text{total}}$, the estimated prevalence of tobacco usage for the kth combination is given by

$$\tilde{p}_k^{\text{user}} = \frac{\widetilde{N}_k^{\text{user}}}{\widetilde{N}_k^{\text{total}}}$$

Now, let k be one of the sex/age group combinations of table 4. Using the same logic, the estimated number of non-users for the kth combination ($k = 1, \ldots, 8$) is given by

$$\widetilde{N}_k^{\text{non-user}} = \sum_{j \in U} w_j^{\text{EHWT}} \, HN_j^k$$

where HN_j^k is the number of enumerated individuals in the jth household that do not use any tobacco product and fall within the sex/age group corresponding to combination k, and the sum is over all enumerated households nationwide (since all 8 provinces were pooled together for estimating non-users). Note: $H_j^k = HU_j^k + HN_j^k$. The formula for $\widetilde{N}_k^{\text{total}}$ remains the same, but the the subscript k now refers

[2]There was also 13 individuals that had both age and tobacco usage status missing. However, to keep the estimation procedure of section 3.2 simple, these individuals were simply excluded. Since there are so few such individuals, doing so has effectively no impact.

to the kth combination of table 4; hence, the sum is now over all enumerated households nationwide. After computing the new $\widetilde{N}_k^{\text{total}}$, the estimated proportion of individuals that do not use any tobacco product for the kth combination is given by

$$\tilde{p}_k^{\text{non-user}} = \frac{\widetilde{N}_k^{\text{non-user}}}{\widetilde{N}_k^{\text{total}}}$$

3.2 With missing values

We now incorporate the 442 + 27 = 469 individuals with missing values. To this end, let HS_j^k be the number of enumerated individuals in the jth household that fall within the sex/age group corresponding to k^{th} combination of table 3, but for whom the tobacco usage status is unknown. In addition, let

$$HA_j^g = \begin{cases} \text{\# of enumerated males in the } j^{\text{th}} \text{ household with missing age} & \text{if } g = 1 \\ \text{\# of enumerated females in the } j^{\text{th}} \text{ household with missing age} & \text{if } g = 2 \end{cases}$$

HA_j^g needs to be divided into those that are tobacco users and those that do not use any tobacco product. Hence, let HAU_j^g be the number of enumerated male ($g = 1$)/female ($g = 2$) tobacco users in the jth household with missing age, and similarly for HAN_j^g; thus, $HAU_j^g + HAN_j^g = HA_j^g$. Let $S_1(g, a, d)$ be the set of all enumerated male ($g = 1$)/female ($g = 2$) respondents in age group a ($a = [15, 25)$, $[25, 35)$, $[35, 50)$ or $[50, 100)$) residing in district d ($d = 1, \ldots, 21$). Let $\theta_d^{a,g}$ be the proportion of those individuals that use tobacco products. In addition, let $S_2(g, d)$ be the set of all enumerated male ($g = 1$)/female ($g = 2$) respondents residing in district d. Let $\beta_d^{(a,g)}$ be the proportion of those individuals that are in age group a; thus, $\sum_a \beta_d^{(a,g)} = 1$. The estimated number of tobacco users for the kth combination ($k = 1, \ldots, 32$) of table 3 is given by

$$\widehat{N}_k^{\text{user}} = \sum_{j \in U_h} w_j^{\text{EHWT}} \Big(HU_j^k + HAU_j^g \ \beta_{d(j)}^{(a,g)} + HS_j^k \ \theta_{d(j)}^{(a,g)} \Big)$$

where $d(j)$ denotes the district where the jth household is located, and (a, g) are the sex/age group corresponding to the kth combination. As with $\widetilde{N}_k^{\text{user}}$, the sum is over all enumerated households in the hth province(s) for males, and over all enumerated households nationwide for females. Similarly, the estimated total number of individuals for the same k^{th} combination is given by

$$\widehat{N}_k^{\text{total}} = \sum_{j \in U_h} w_j^{\text{EHWT}} \Big(H_j^k + HA_j^g \ \beta_{d(j)}^{(a,g)} + HS_j^k \Big)$$

Combining $\widehat{N}_k^{\text{user}}$ and $\widehat{N}_k^{\text{user}}$, the estimated prevalence of tobacco usage for the kth combination of table 3 is given by

$$\hat{p}_k^{\text{user}} = \frac{\widehat{N}_k^{\text{user}}}{\widehat{N}_k^{\text{total}}}$$

Now, let k be one of the sex/age group combinations of table 4. Using the same logic, the estimated number of non-users for the kth combination ($k = 1, \ldots, 8$) is given by

$$\widehat{N}_k^{\text{non-user}} = \sum_{j \in U} w_j^{\text{EHWT}} \Big(HN_j^k + HAN_j^g \ \beta_{d(j)}^{(a,g)} + HS_j^k \ (1 - \theta_{d(j)}^{(a,g)}) \Big)$$

As with $\tilde{p}_k^{\text{user}}$, the formula for $\widehat{N}_k^{\text{total}}$ remains the same, but the subscript k now refers to the kth combination of table 4; hence, the sum is now over all enumerated households nationwide. After computing the new $\widehat{N}_k^{\text{total}}$, the estimated proportion of individuals that do not use any tobacco product for the kth combination of table 4 is given by

$$\hat{p}_k^{\text{non-user}} = \frac{\widehat{N}_k^{\text{non-user}}}{\widehat{N}_k^{\text{total}}}$$

k	sex	province(s)	age	$\hat{p}_k^{\text{user}}$	$\widehat{N}_k^{\text{user}}$
1	male	Nairobi	[15-24)	13.3%	32,660
2	male	Nairobi	[25-34)	27.9%	126,564
3	male	Nairobi	[35-49)	28.9%	80,005
4	male	Nairobi	[50-100)	37.1%	27,115
5	male	Central	[25-34)	43.3%	128,117
6	male	Central	[35-49)	41.2%	147,073
7	male	Central	[50-100)	42.1%	173,347
8	male	Coast	[15-24)	35.0%	115,919
9	male	Coast	[25-34)	49.1%	197,905
10	male	Coast	[35-49)	45.1%	123,804
11	male	Eastern	[15-24)	15.6%	80,619
12	male	Eastern	[25-34)	37.8%	152,534
13	male	Eastern	[35-49)	44.2%	162,705
14	male	Eastern	[50-100)	55.6%	223,857
15	male	North Eastern	[15-24)	17.1%	3,561
16	male	North Eastern	[25-49)	26.2%	7,252
17	male	North Eastern	[50-100)	37.8%	3,324
18	male	Nyanza	[25-34)	18.0%	67,127
19	male	Nyanza	[35-49)	25.2%	74,994
20	male	Nyanza	[50-100)	27.9%	86,469
21	male	Rift Valley	[15-24)	10.2%	89,150
22	male	Rift Valley	[25-34)	28.3%	256,306
23	male	Rift Valley	[35-49)	33.7%	206,973
24	male	Rift Valley	[50-100)	40.8%	165,554
25	male	Western	[25-34)	18.9%	51,447
26	male	Western	[35-49)	26.7%	68,511
27	male	Coast+Western	[50-100)	23.8%	74,664
28	male	Central+Nyanza+Western	[15-24)	3.8%	17,473
29	female	all	[15-24)	0.9%	31,758
30	female	all	[25-34)	3.2%	89,437
31	female	all	[35-49)	4.2%	90,049
32	female	all	[50-100)	19.9%	362,613

Table 3: Estimated numbers of individuals that use one or more tobacco products, per sex/province(s)/age group for males and sex/age group for females, and corresponding prevalences, at the time of Wave 1.

k	sex	age	$\hat{p}_k^{\text{non-user}}$	$\widehat{N}_k^{\text{non-user}}$
1	male	[15-24)	88.4%	2,789,174
2	male	[25-34)	68.7%	2,140,830
3	male	[35-49)	64.7%	1,587,803
4	male	[50-100)	60.8%	1,171,712
5	female	[15-24)	99.1%	3,567,341
6	female	[25-34)	96.8%	2,691,307
7	female	[35-49)	95.8%	2,038,155
8	female	[50-100)	80.1%	1,463,146

Table 4: Estimated numbers of individuals that do not use any tobacco product, per sex/age group, and corresponding prevalences, at the time of Wave 1.

3.3 Estimated prevalences and targets for the time of Wave 2

Table 5 gives the estimated prevalences and targets at the time of Wave 2 for tobacco users; Table 6 gives the estimated prevalences and targets at the time of Wave 2 for non-users.

k	sex	age	$\hat{p}_k^{\text{user}}$	$\widehat{N}_k^{\text{user}}$	
1	male	all but North Eastern	[15-29)	17.0%	1,033,145
2	male	Nairobi+Rift Valley	[30-44)	21.7%	312,755
3	male	Nairobi+Rift Valley	[45-59)	27.8%	179,158
4	male	Nairobi+Rift Valley	[60-100)	36.3%	150,710
5	male	Central	[30-44)	36.0%	157,609
6	male	Central	[45-59)	34.5%	112,718
7	male	Central	[60-100)	31.6%	74,481
8	male	Coast+Eastern	[30-44)	39.1%	326,465
9	male	Coast+Eastern	[45-59)	35.4%	154,334
10	male	Coast+Western	[60-100)	17.4%	25,495
11	male	Eastern	[60-100)	49.0%	104,172
12	male	North Eastern	all	29.7%	14,137
13	male	Nyanza	[30-44)	12.4%	53,780
14	male	Nyanza	[45-59)	25.1%	66,683
15	male	Nyanza	[60-100)	16.2%	22,671
16	male	Western	[30-44)	18.3%	62,683
17	male	Western	[45-59)	15.3%	27,648
18	female	all	[15-29)	2.80%	176,588
19	female	all	[30-44)	3.30%	123,148
20	female	all	[45-59)	7.30%	124,415
21	female	all	[60-100)	10.20%	110,749

Table 5: Estimated numbers of individuals that use one or more tobacco products, per sex/province(s)/age group for males and sex/age group for females, and corresponding prevalences.

For Tables 5 and 6 the population estimates, by age/gender and as of 2014, were obtained from the 2015 Kenya Facts and Figures published by the Kenya National Bureau of Statistics. The prevalence

k	sex	age	$\hat{p}_k^{\text{non-user}}$	$\widehat{N}_k^{\text{non-user}}$
1	male	[15-29)	83.0%	5,054,981
2	male	[30-44)	73.6%	2,627,100
3	male	[45-59)	70.9%	1,193,633
4	male	[60-100)	66.4%	636,317
5	female	[15-29)	97.2%	6,130,126
6	female	[30-44)	96.7%	3,608,595
7	female	[45-59)	92.7%	1,579,905
8	female	[60-100)	89.8%	975,021

Table 6: Estimated numbers of individuals that do not use any tobacco product, per sex/age group, and corresponding prevalences, at the time of Wave 2.

estimates, by age/gender, for the use of tobacco products were obtained from the 2015 Kenya STEPwise Survey published by the Kenya National Bureau of Statistics and Ministry of Health.

Acknowledgements

Core funding for the ITC Project is provided by the U.S. National Cancer Institute (NCI; grants P50 CA111236 & P01 CA138389), the Canadian Institutes of Health Research (CIHR; grants 79551 & 115016), and by the Ontario Institute for Cancer Research (OICR).

References

[1] Hanurav, T. V. (1967) Optimum Utilization of Auxiliary Information: π_{ps} Sampling of Two Units from a Stratum. *Journal of the Royal Statistical Society, Series B* 29, 374–391.

[2] Kenya National Bureau of Statistics (KNBS) & ICF Macro (2010) *2008-09 Kenya Demographic and Health Survey.* Calverton, Maryland.

[3] Kenya National Bureau of Statistics (KNBS)(2010) *2009 Kenya Population and Housing Census.* Nairobi, Kenya.

[4] Levy, P. S. & Lemeshow, S. (2008) *Sampling of Populations: Methods and Applications*, 4 edn., John Wiley & Sons, Hoboken, NJ.

[5] Little, R. J. A. & Rubin, D. B. (2002) *Statistical Analysis with Missing Data*, 2 edn., John Wiley & Sons, Hoboken, NJ.

[6] Murthy, M. N. (1957) Ordered and Unordered Estimators in Sampling without Replacement. *Sankhyā* 18, 379–390.

[7] Patton, M. Q. (1990) *Qualitative Evaluation and Research Methods.* 2nd edition. Sage, Newbury Park, CA.

[8] Vijayan, K. (1968) An Exact π_{ps} Sampling Scheme: Generalization of a Method of Hanurav. *Journal of the Royal Statistical Society, Series B* 30, 556–566.